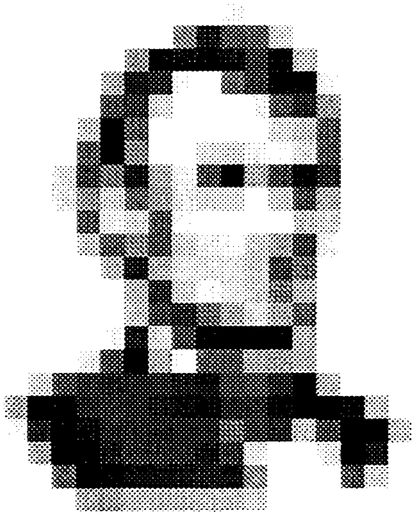
original
/3/99

Speech Processors for Auditory Prostheses

NIH Contract N01-DC-92100

QPR #1: Jan-Feb-Mar 1999



Robert V. Shannon
Fan-Gang Zeng
Qian-Jie Fu
Monita Chatterjee
John Wygonski
John Galvin III
Mark Robert
Xiaosong Wang

House Ear Institute
2100 W. Third St.
Los Angeles, CA 90057

26 April 1999

Abstract

In this first progress report, we describe the development of a research interface for the Clarion cochlear implant and the initial planning for the design of a research interface for the Nucleus 24M cochlear implant system. Three experiments are described that are in progress: electrode interaction, effect of number of channels on speech recognition in noise, and the effect of uneven neuronal survival on speech recognition. Electrode interaction is being measured with forward masking as a function of the pulse duration, electrode stimulation mode and pulse rate. A simple model is proposed to describe the interaction data and relate it to the growth of loudness. Measures of speech and phoneme recognition are being collected from Clarion and Nucleus implant patients as a function of the number of electrodes used in their speech processor. These measures are being collected in quiet and as a function of the signal-to-noise level. Speech and phoneme recognition are also being measured in normal-hearing listeners and in cochlear implant listeners under conditions of "holes" in hearing. Experimental conditions are constructed to simulate a region of missing neurons at the apex, middle or base of the cochlea. The speech information that would normally be delivered to the hole region is delivered either to the apical or basal edge of the hole or is simply dropped. Three manuscripts are included in the appendix of recently completed studies.

Introduction

This is the first progress report of a new contract to develop speech processors for electrically stimulated prosthetic hearing. The work scope of the contract contains 3 main areas of work: basic psychophysics and physiology of electrical stimulation relevant to speech processor design, speech processor parameter assessment, and construction of wearable speech processors to assess the impact of long-term learning on speech processor performance. The quarterly progress reports from this contract will generally contain two parts: a brief overview of all work in progress, and a formal description of a completed study.

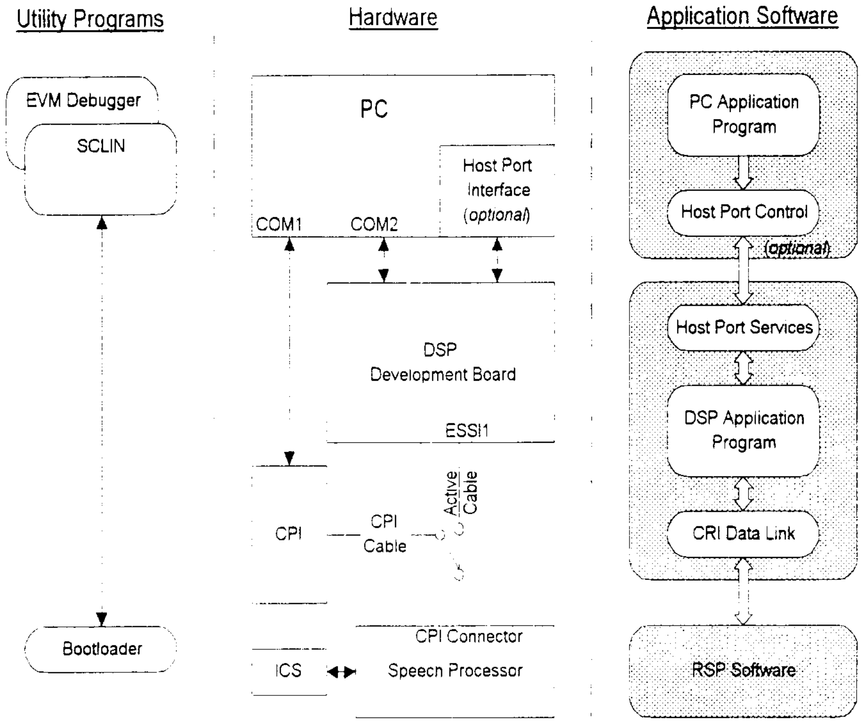
In this first report, we describe our first three months of work. Most of the projects funded by this contract have a history of previous work funded internally by the House Ear Institute as pilot projects. The progress reported in this report reflects the completion of some ongoing projects that were funded by the House Ear Institute but are now funded by this contract.

Implant Interface Development

Clarion Research Interface

House Ear Institute developed a research interface to the Clarion cochlear implant under a joint development agreement with Advanced Bionics Corporation in late 1997 and 1998. The Clarion Research Interface (CRI) is intended for use by technically sophisticated research groups who desire to work with Clarion cochlear implant subjects. [*Note: A workshop was held in Sylmar, California on April 16-18, 1999 to train representatives of implant research groups in the use of the CRI. Representatives of 9 research groups from the US attended: MIT, RTI, Duke, Indiana, Minnesota, UT Dallas, Arkansas, Iowa, and the House Ear Institute. At the conclusion of the 2.5-day workshop each group was fully trained in the operation of the new interface hardware and software.*]

Use of the interface requires programming of real-time digital signal processor (DSP) chips. The interface will allow (1) the presentation of preprocessed stimuli, (2) or presentation of speech files from a host PC computer, or (3) the implementation of real-time speech processors. In each case it is the experimenters' responsibility to ensure that the stimuli are appropriate for the dynamic range of the particular subject. The following figure presents an overview of the system architecture of the CRI.



CRI Hardware

The hardware consists of a host PC capable of running Clarion SCLIN software, a DSP Development Board (EVM), and the Clarion Speech Processor(SP)/Headpiece and Implantable Cochlear Stimulator (ICS). Utility programs have been developed to configure the SP with the research interface software and to configure the DSP Development Board with application programs for stimulus generation. DSP Demo Application software serves as a model for constructing user applications for speech processor strategy development and psychophysical testing. An optional host port interface ISA card enables PC interaction with DSP programs and high-speed transfer of data between DSP and PC.

The hardware components required for the Clarion Research Interface are:

1. Host PC. The host PC is a standard PC computer running Windows. The capabilities of this PC are the same as those required to run SCLIN.
2. EVM DSP Board. The DSP board is a DSP 56302 Evaluation Module (DSP56302EVM) from Motorola.
3. Host Port Interface. ISA Host Interface for DSP 5630xEVM from Domain Technologies.
4. CPI. Clinician's Programming Interface, part of the clinical equipment provided with the Clarion system.
5. SP/HP. Clarion Speech Processor and Headpiece.
6. RICS. The Research ICS (RICS) is a specially-designed ICS that allows probing of signals at the ICS outputs which are connected to load

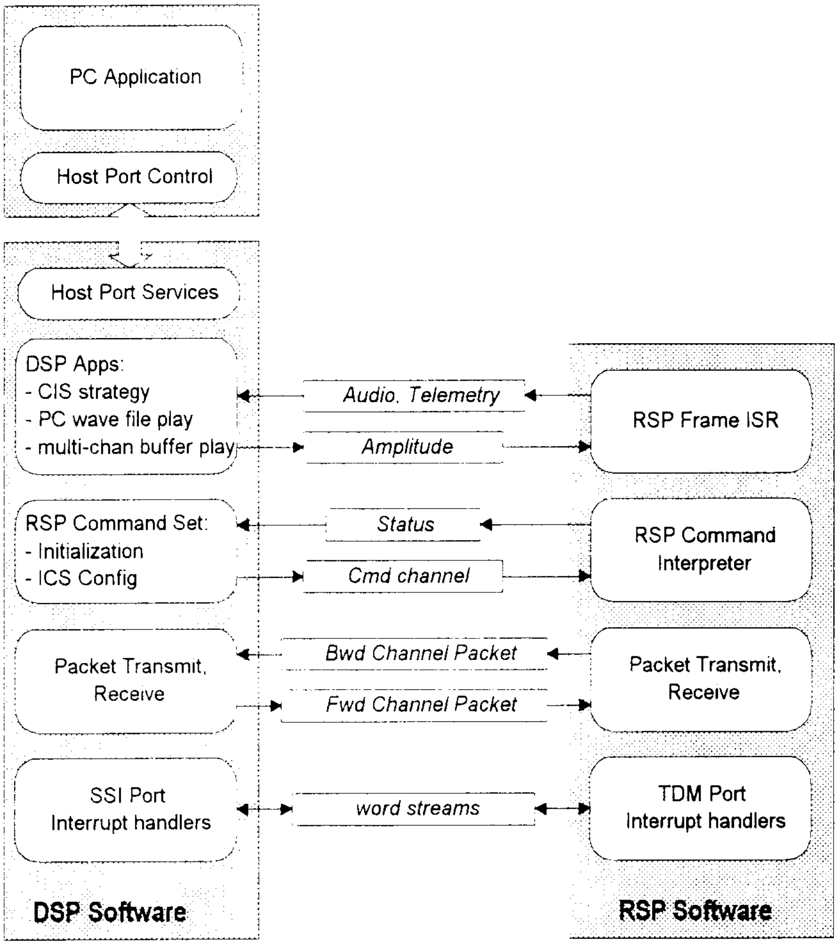
resistors. The RICS is used in place of the subject's ICS when in development mode.

7. SP/DSP Cable. The DSP ESSI1 port available at J6 on the DSP56302 EVM is connected to the CPI connector on the SP with a special cable having active drive circuitry for high-speed serial port signals.

8. Power supply. Power to the RSP is supplied by the DSP56302EVM over the SP/DSP cable used for data link communications. The DSP56302EVM is powered by a 7-9V, 1A AC or DC supply with 2.1mm plug, which supplies power for the EVM, SP, and active cable.

CRI Software

Software for the Clarion Research Interface consists of RSP software running on the SP and DSP software running on the EVM. In addition, an application may require custom software running on the PC to interact with the EVM through the Host Port Interface. The architecture of CRI software is shown conceptually in the diagram below.



The software provided with the CRI consists of DSP programs running on the EVM that demonstrate the basic capabilities of the Clarion Research

Interface. The DSP programs may also require interaction with the PC through the Host Port Interface, depending on the application. The RSP software consists of a single program downloaded to the SP by the researcher.

The Demo software provided with the interface consists of three application programs that demonstrate the functionality of the Clarion Research Interface. These programs are provided as examples and the user will typically want to modify and enhance these programs to support their research needs.

Demo 1: Multiple-channel waveform array The DSPDEMO.ABS program executes on the EVM and demonstrates the capability of the Clarion Research Interface to generate arbitrary, periodic waveforms at ICS outputs. The DSP simply manages the transfer of ICS amplitudes for 8 channels that are contained in a buffer in the data memory of the DSP. The data is an array of initialized constants comprising appropriate charge-balanced stimuli.

Demo 2: CIS Strategy using audio from HP microphone The DSPCIS demonstration program is a simplified CIS speech processing strategy that runs in real-time on the DSP board. The DSP program generates ICS outputs every frame and passes them over the forward channel of the DSP-SP communications link, resulting in waveforms generated at the electrodes. The source of the input data for the CIS algorithm is digitized audio words from the HP microphone that are sent to the DSP, one word every frame, over the backward channel of the DSP-SP communications link.

Demo 3: PC ".wav" file play through CIS Strategy This demonstration involves a PC program sending data to a program running on the EVM through the host port interface for real-time processing. The DSPCISW DSP program that executes on the EVM is nearly identical in function to DSPCIS except that the source of the audio input to the CIS strategy is a stream of audio samples from the PC. A typical use of this method is for playing arbitrarily-long digitized test materials such as word lists and sentences through the Clarion system. The PC program implements the '.wav' file play and the host port interface functions on the PC side.

NOTE: The Clarion Research Interface is available to qualified research groups through the Advanced Bionics Corporation.

Nucleus 24M Interface.

Our present experiments with Nucleus-22 cochlear implant subjects are implemented with a custom designed interface that allows presentation of arbitrary stimuli to any electrode combination (Shannon et al., 1990). The new system now in clinical use, the Nucleus-24M system, uses a different signal transmission protocol than the Nucleus-22 system. To achieve higher throughput pulse rates the pulse parameter information for one pulse is transmitted simultaneously with the previous pulse waveform.

We have received technical specifications from Cochlear Corp. on the design of their signal transmission system and are evaluating the best method for developing a research interface for the 24M device.

Overview of Experiments in Progress

Channel Interaction (work partially funded by contract)

One of the main factors thought to be of importance for cochlear implant function is the degree of interaction or overlap between electrodes. Electrodes that stimulate a local region will allow good tonotopic selectivity, while electrodes that stimulate a broad region will only allow a blurred or smeared representation of the same spectral information. Our group has developed several methods for measuring electrode interaction. Monita Chatterjee has measured the spread of activity around a stimulated electrode using forward masking. The amount of forward masking produced by a stimulus is an indicator of the excitation pattern it evokes. In this method, a 200 ms masking stimulus is presented to a fixed electrode pair. The masker was varied in its level, electrode mode and pulse duration. Five msec after the masker is turned off, a brief (20 ms) probe stimulus is presented on another electrode pair. The probe was always 500 Hz, 200 µs/phase biphasic pulse in stimulation mode BP+1. The amplitude of the probe is adjusted to the point where it is just detectable. This threshold is then measured as a function of the separation between masker and probe location as an indication of the spread of excitation of the masker. Stimuli were presented using a custom implant interface (Shannon et al., 1990). Experiments were run using custom software designed by Q.J.Fu. Subjects were three highly trained users of the Nucleus-22 implant system.

Effect of increasing masker stimulation mode (separation between electrodes). Results are plotted as linear threshold shifts (masked – quiet threshold) in microamps (µA) vs. probe electrode location. Figure 3 shows examples of masked patterns obtained in subjects N3, N4 and N7. Masker pulse duration was fixed at 200 microseconds/phase. Three sets of data are shown, each for a different masker electrode pair. Upper and lower panels of each set show masked patterns obtained using masker amplitudes at 50% and 70% of the dynamic range of the masker, respectively. Typically, on a subjective loudness scale of 0 to 100, stimuli presented at 50% of the dynamic range are judged to be at a loudness of about 20 (soft), and stimuli at 70% of the dynamic range are judged to be at a loudness of about 35 (comfortable). Within each panel, each symbol corresponds to data obtained in a different subject. In general, the pattern becomes more broad as masker level increases and with larger separation between masker electrodes.

Effects of increasing masker pulse duration. Examples of results obtained in subject N7 are shown in Fig. 4. In this case, the masker electrode pair is fixed, and the masked pattern is obtained for different masker pulse durations. For each pulse duration of the masker, the masking pattern is obtained for different masker pulse amplitudes. Note that changing the masker pulse duration does not change the overall shape of the excitation pattern in any significant manner.

Channel Interaction – DISCUSSION. As the separation between the electrodes of the masker electrode pair increases, the pattern of masked threshold shift broadens. At very large separations (such as masker electrode pair 2,20), the excitation pattern resembles that due to two monopoles. As the electrode array does not extend beyond 22 electrodes, we cannot measure the threshold shift (or the excitation) pattern completely under these conditions. Interestingly, masker pulse duration does not seem to influence the gross shape of the excitation pattern.

In general, it is known that loudness grows slowly for small electrode separations (narrow configurations) and rapidly for large electrode separations. For a given electrode separation, loudness grows slowly for short pulses, and rapidly for long pulses.

In our previous studies, we have quantified these relations. We have found that for simple stimuli such as these, loudness grows as a function of electrode separation, as well as pulse duration and pulse amplitude, according to an exponential function, the exponent α of which is given by: $\alpha = g(M)*f(D)*I$ where g(M) is a linear function of the electrode separation M (M = 2 for a stimulation mode of BP+1) and f(D) is a compressive power function of pulse duration D, I being the stimulus amplitude.

Making a broad simplification, we explore the premise that the forward masked pattern may be thought of as an "excitation pattern", and the area under it may ideally represent some form of integrated neural activity due to the stimulus. If loudness is monotonically related to such an integrated output, the area under the pattern should behave in a qualitatively similar manner to subjective loudness in these implant listeners.

Figure 5 shows the area under the excitation pattern plotted against masker amplitude, for three masker electrode pairs (narrowest to broadest separation, from top to bottom). In each case, masker pulse duration is fixed at 200 microseconds/phase. In each panel, different symbols correspond to a different subject (indicated).

Note that the area under the pattern grows most shallowly for the smallest masker-electrode separation, and most rapidly for the widest separation. The function relating the growth of the area to current amplitude is approximately linear, and very similar across subjects. These relations parallel surprisingly well the behavior of the loudness functions.

Figure 6 shows the slope of the area-vs.-amplitude function plotted against the electrode separation, for subject N7. An approximately linear behavior is observed. Note that due to the limited length of the electrode array, we are underestimating the area of activation for the widest masker-electrode separations.

Figure 7 plots the area under the excitation pattern for a fixed masker electrode pair (10,12) in subject N7. Different symbols correspond to different masker pulse durations. It is apparent that for the shortest pulse duration, the area grows with the shallowest slope, while for the longest pulse duration, the area grows with the steepest slope. Again, this behavior parallels the behavior of the loudness function.

Figure 8 plots the slope of the area-vs.-amplitude functions against pulse duration, for two different masker electrode separations. Note that the slope increases with pulse duration. in a manner that can be approximated well by a compressive power function. Note that the best fitting power functions for the two masker configurations have almost identical exponents (0.39). This behavior is again similar to that of the loudness function. The value of the exponent of the power function relating the slope of the loudness function to the pulse duration is approximately 0.6, somewhat larger than the 0.39 observed here.

The results described above suggest that the area under the excitation pattern is monotonically related to perceived loudness. Given that the area under the pattern is poorly estimated because of the limitations already described, these findings are both interesting and satisfying.

Number of channels in noise

Several studies have investigated the effect of the number of spectral channels (or number of electrodes) on speech recognition (Fishman et al., 1997; Fu and Shannon, 1998). As might be expected. speech recognition improves as the number of spectral channels increases. However. in cochlear implant listeners, increasing the number of electrodes only improved speech recognition performance up to 7 electrodes. There was no improvement as the number of electrodes was increased from 7 to 20. However. Fishman et al. measured performance in quiet. Fu and Shannon (1998) showed that more spectral channels are necessary when listening in noisy conditions. In the most difficult noise conditions, speech recognition performance had not reached asymptote at 16 channels of spectral information. It is possible that performance will improve from 7 to 20 electrodes in noisy listening conditions. though Fishman et al. found no improvement in quiet. We are measuring speech recognition in 10 Nucleus-22 subjects and in 10 Clarion subjects at 5 noise levels with varying numbers of electrodes. Results will be reported at the upcoming Asilomar CIAP meeting August 29-September 3. 1999.

Holes in hearing

The remaining neuron population in a cochlear implant subject may be quite variable across individuals due to the cause of deafness and its duration. Some pathologies causing deafness may have an uneven effect on nerve survival, resulting in patches of cochlear with no remaining neurons. A cochlear implant in such a case would be able to produce sound sensations, but the current level on the electrode in the "dead zone" would have to be increased so that the electrical field spreads to excite neurons on the edge of the dead zone. In such a case the speech information carried by that electrode would still be represented in the neural stimulation, but it would occur at the wrong cochlear location. We are conducting a series of experiments to quantify the effect of such "holes" in hearing. We are systematically creating holes in the tonotopic representation of cochlear implants by turning off one or more electrodes. The speech information that had been applied to that electrode could then be routed to adjacent electrodes to simulate the condition described above. Alternatively, we can simply drop the speech information that had been assigned to that electrode. We are presently creating "holes" at the apical, middle and basal regions in Nucleus-22 subjects. Speech recognition is being measured in conditions which turn off 2, 4, 6, or 8 electrodes resulting in holes of 1.5, 3.0, 4.5 and 6.0 mm, respectively. The speech information that would have normally been assigned to these electrodes is either (1) dropped, (2) routed to electrodes on the apical edge of the hole, (3) routed to electrodes on the basal edge of the hole, or (4) split evenly between the apical and basal edge of the hole. Results will be reported at the upcoming Asilomar CIAP meeting August 29-September 3, 1999.

Long-Term Learning Effects

One issue in designing implant speech processing strategies is the influence of long-term learning. It is widely thought that patients will "get used to" or accommodate to many processor parameters even if they are not optimal. Section D of the workscope for this contract calls for the "study of the effects of learning". We are presently engaged in a study of adaptation to a speech processor that is purposefully mismatched in terms of frequency-to-electrode assignments. We have three Nucleus-22 implant volunteers who are wearing a processor in which the frequency-to-electrode assignment has been shifted by 3 mm in cochlear distance. Speech recognition measures are made weekly during the trial. Results will be reported at the upcoming Asilomar CIAP meeting August 29-September 3, 1999.

Anticipated Work in the Next Quarter

In the next quarter we will continue hardware and software development of interfaces for the Nucleus 24 and Clarion implant systems. PC programs are in development to allow general research control of stimulation of the Clarion system. We anticipate that these programs will be completed in the next quarter and will be available for use. Also in the next quarter we will formalize the hardware/software approach that we will use to develop a research interface for the Nucleus 24M implant system and work will begin on the interface development.

We anticipate that the four experiments briefly described in the "In Progress" section above will be completed and manuscriopts submitted for publication. New experiments will be initiated on (1) the effects of noise on speech recognition, (2) long-term learning effects, (3) interleaved masking as a measure of electrode interaction, (4) further studies on the effect of stimulation pulse rate on speech recognition, and (5) comparison of speech recognition with different stimulation modes.

Summary of Experiments in Appendix

Nonlinear Amplitude Mapping in Noise: Acoustic and Electric Hearing

Two of the manuscripts in the Appendix present data on the effects of various amplitude mapping functions on phoneme recognition in noisy listening conditions. Five normal-hearing listeners were tested with a sixteen-band noise vocoder to restrict their spectral resolution. Three Nucleus-22 implant patients were tested with a 4-band CIS speech processor. In both cases an instantaneous power-law was used to map envelope amplitudes in each band to a new amplitude that was used to modulate the carrier signal: a noise band in the acoustic case, and 500-Hz biphasic pulses in the implant case. The exponent of the power-law mapping was varied over a range from more compressive to more expansive than the exponent that produced "normal" loudness growth. Thus, normal-hearing listeners were tested with exponents ranging from 0.3 to 3.0, where an exponent of 1.0 would preserve normal loudness growth. Implant listeners were tested with exponents ranging from 0.05 to 0.8 (where normal loudness growth was achieved by an exponent of 0.2).

In quiet listening conditions only a small decrement in vowel and consonant recognition was observed over the entire range of exponents. A mild peak in recognition performance was observed with the exponent that preserved the normal loudness. However, in noisy listening conditions the performance functions become clearly asymmetric. For both normal-hearing and cochlear implant listeners, and for both vowels and consonants, performance drops dramatically with increasing noise level for compressive exponents but drops only slightly for expansive exponents. With signal-to-noise ratios of 0 dB or poorer the best performance was observed with the most expansive exponent used. These results confirm that preserving normal loudness growth produces the best phoneme recognition in quiet. However, in noise, the best amplitude mapping function is more expansive than one that preserves normal loudness growth. The fact that the functions were quite similar in shape for both NH and CI listeners indicates that this result is a general one.

For expansive exponents the peaks in the waveform in each channel are accentuated. The reason performance is better in this case is that the peaks are more likely to be from the speech than from the noise, so the expansive amplitude transformation is differentially amplifying peak amplitude samples in speech relative to the noise. This suggests an extremely simple preprocessing strategy for listening in noise – expansive instantaneous amplitude mapping.

Phoneme Recognition as a Function of Stimulation Rate.

One of the critical parameters in cochlear implant speech processor design is the rate of stimulation on each electrode. We measured vowel and consonant recognition in three Nucleus-22 implant patients as we systematically changed the stimulation pulse rate on each electrode. For this experiment we implemented a four-channel CIS processor on the Nucleus-22 system using pre-processed stimuli. Four BP+3 electrode pairs were selected for stimulation that covered most of the electrode array: (18.22). (13,17), (8,12), and (3,7). The stimulation rate/channel was varied from 50 pps to 500 pps. Results showed that phoneme recognition performance increased with pulse rate up to 150 pps, and no further increases were observed as pulse rate was increased from 150 pps to 500 pps/channel. An additional condition confirmed that the asymptote at 150 pps was not due to the envelope smoothing filter. Thus, this experiment demonstrated that, within the range of 50-500 pps/electrode, no improvement in phoneme recognition was obtained for stimulation rates higher than 150 pps. This asymptotic rate is considerably lower than stimulation rates used by present implant systems.

References

Fishman K., Shannon R.V. and Slattery W.H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. **Journal of Speech and Hearing Research**, 40, 1201-1215.

Fu, Q.-J., Shannon, R.V., and Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing, **Journal of the Acoustical Society of America**, 104(6), 3586-3596.

Shannon, R.V., Adams, D.D., Ferrel, R.L.. Palumbo, R.L., and Grandgenett, M. (1990). A computer interface for psychophysical and speech research with the Nucleus cochlear implant. **J. Acoust. Soc. Am.** 87, 905-907.

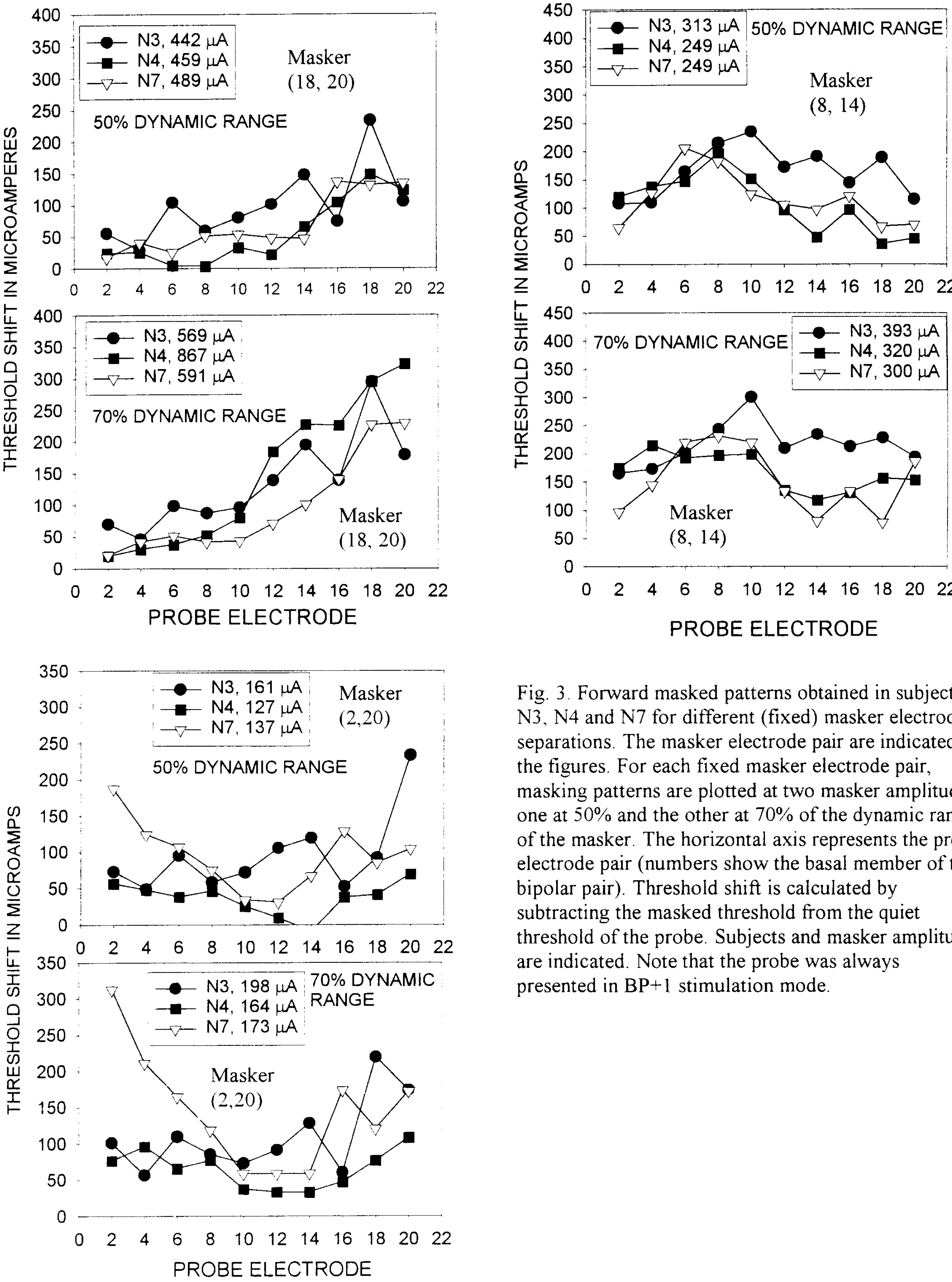


Fig. 3. Forward masked patterns obtained in subjects N3, N4 and N7 for different (fixed) masker electrode separations. The masker electrode pair are indicated in the figures. For each fixed masker electrode pair, masking patterns are plotted at two masker amplitudes one at 50% and the other at 70% of the dynamic range of the masker. The horizontal axis represents the prob electrode pair (numbers show the basal member of the bipolar pair). Threshold shift is calculated by subtracting the masked threshold from the quiet threshold of the probe. Subjects and masker amplitude are indicated. Note that the probe was always presented in BP+1 stimulation mode.

SUBJECT: N7, MASKER (8,14)

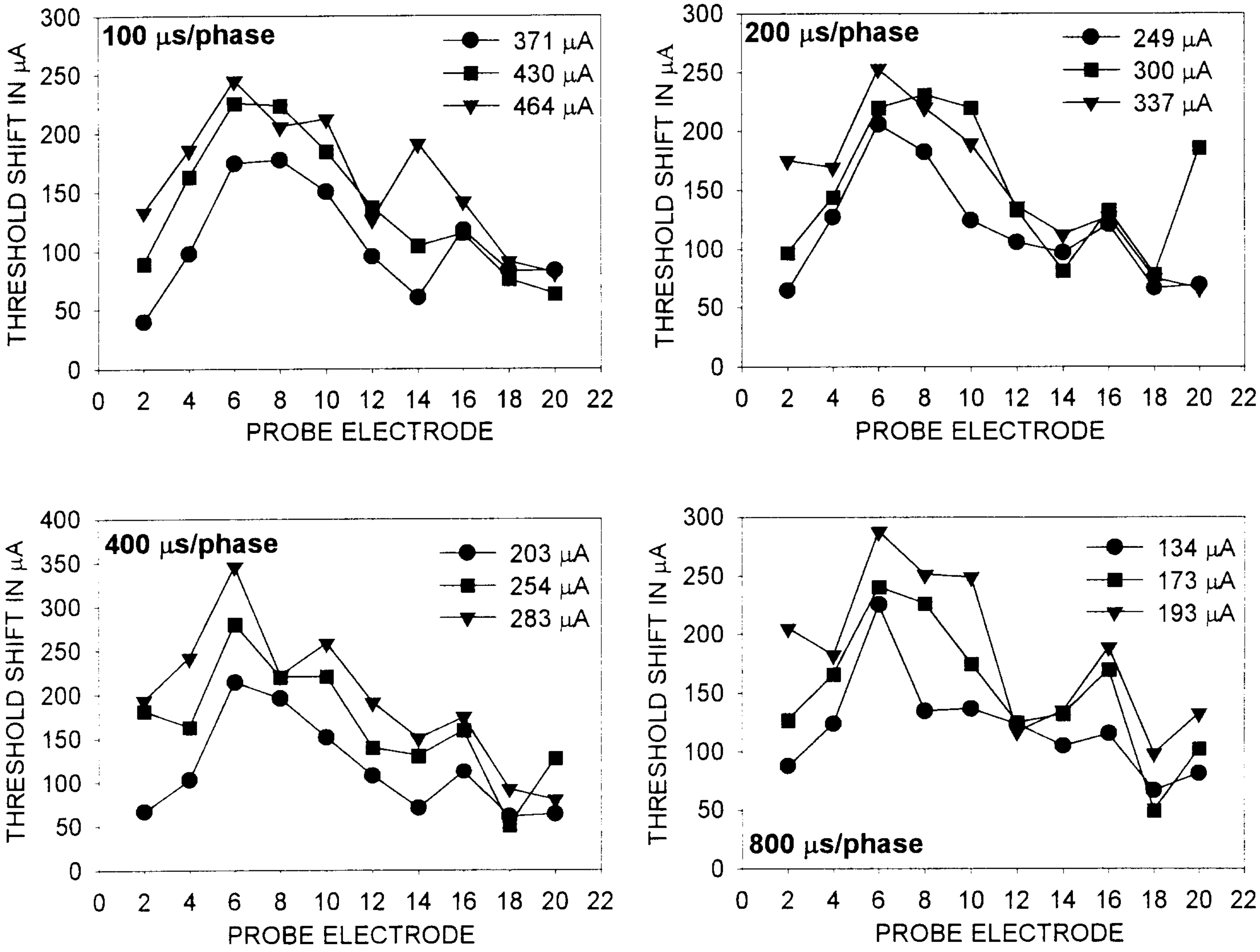


Fig. 4 Masked patterns obtained in subject N7. Different panels correspond to data obtained using different (fixed) masker pulse durations (indicated in each panel). All data were obtained using a fixed masker electrode pair (8, 14), corresponding to a BP+5 stimulation mode. Probe was always presented in BP+1 stimualtion mode. Within each panel, different symbols correspond to data obtained using different masker amplitudes (at 50%, 70% and 80% of the dynamic range of the masker for the relevant pulse duration).

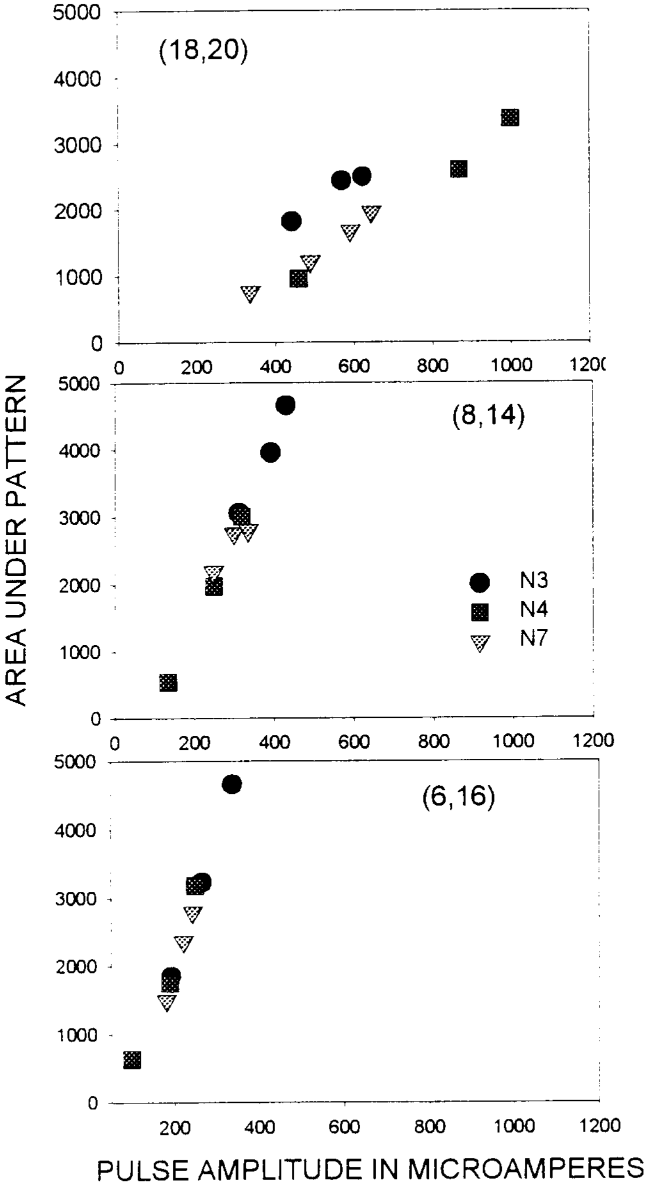


Figure 5. Area under the excitation patterns plotted against masker pulse amplitude. In all cases, masker pulse duration is fixed at 200 microseconds/phase. Masker electrode pairs are indicated in the panels. Each symbol in each panel corresponds to data from a different subject.

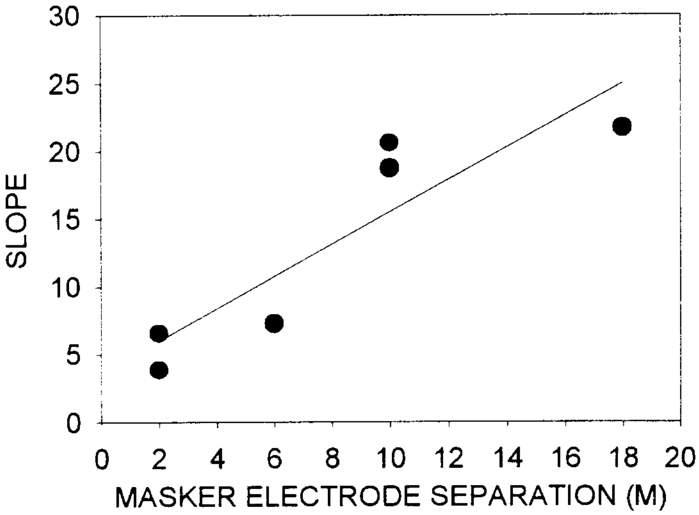


Fig. 6. The slope of the area-vs.-masker amplitude functions (such as those at left) plotted against masker electrode separation, reveals a monotonic, perhaps linear, function. Subject: N7.

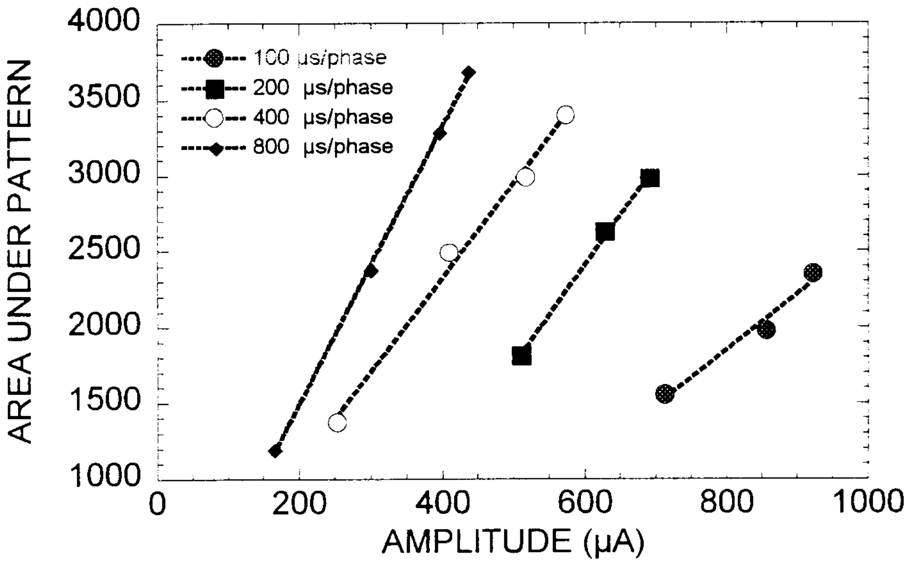


FIG. 7 The areas under the forward masked patterns (subject N7) grow linearly with masker amplitude. The parameter is masker pulse duration (D). Masker electrode pair fixed at (10, 12).

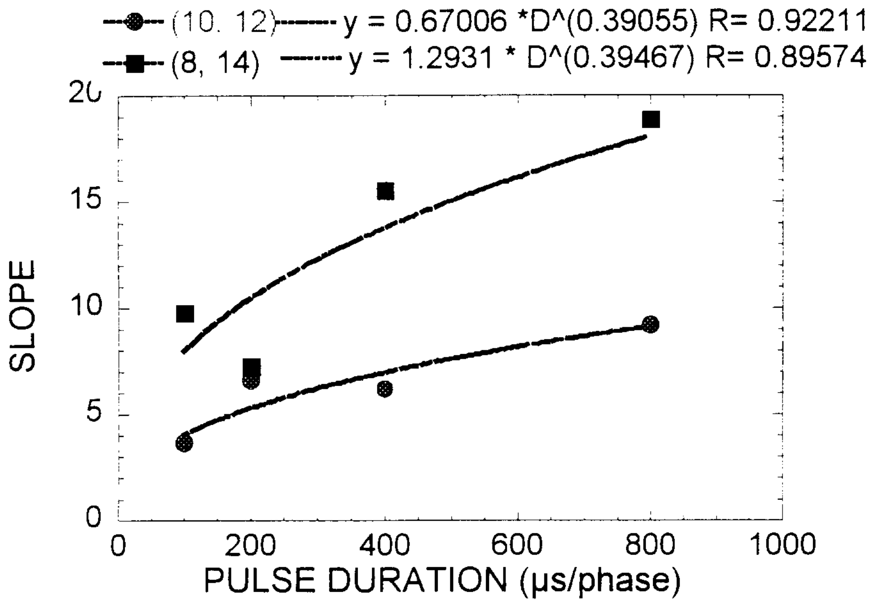


FIG. 8. The slopes of the area-vs.-amplitude functions above are plotted against the masker pulse duration. The symbols correspond to two masker electrode pairs: (10, 12) and (8, 14). The dashed lines represent power function curve fits to the data (equations shown above the figure).

Appendix 1: Completed Experiments: Manuscripts

Fu, Q.-J., and Shannon, R.V. (1999a). Recognition of spectrally degraded speech in noise with nonlinear amplitude mapping, Proceedings of the 1999 IEEE Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 369-372.

Fu, Q.-J, and Shannon, R.V. (1999b). Phoneme recognition as a function of signal-to-noise ratio under nonlinear amplitude mapping by cochlear implant users, Journal of Acoustic Research Online (JASA online), submitted April 99.

Fu, Q.-J. and Shannon, R.V. (1999c). Effect of stimulation rate on phoneme recognition in cochlear implants, Journal of the Acoustical Society of America, (Submitted 10 Dec 98).

RECOGNITION OF SPECTRALLY DEGRADED SPEECH IN NOISE WITH NONLINEAR AMPLITUDE MAPPING

Qian-Jie Fu and Robert V. Shannon

Department of Auditory Implants and Perception
House Ear Institute, 2100 West Third Street
Los Angeles, CA 90057

ABSTRACT

The present study measured phoneme recognition as a function of signal-to-noise level under conditions of spectral smearing and nonlinear amplitude mapping. Speech sounds were divided into 16 analysis bands. The envelope was extracted from each band by half-wave rectification and low-pass filtering and was then distorted by a power-law transformation whose exponents varied from a strongly compressive (p=0.3) to a strongly expanded value (p=3.0). This distorted envelope was used to modulate a noise which was spectrally limited by the same analysis filters. Results showed that phoneme recognition scores in quiet were reduced only slightly with either expanded or compressed amplitude mapping. As the level of background noise was increased, performance deteriorated more rapidly for both compressed and linear mapping than for the expanded mapping. These results indicate that, although an expansive amplitude mapping may slightly reduce performance in quiet, it may be beneficial in noisy listening conditions.

1. INTRODUCTION

Cochlear implants transform speech sounds into electrical signals that directly stimulate remaining auditory nerve fibers and can partially restore the speech sensations of profoundly deaf listeners. Modern multichannel cochlear implants divide speech sounds into multiple frequency bands and extract the temporal envelope information from each band. Then the acoustic envelope amplitude is converted into electric amplitude which is delivered to electrodes located in the different places within the cochlea. To recreate the tonotopic distribution of activity within the normal cochlea, the envelopes from low frequency bands are delivered to electrodes located near the apex and the envelopes from high frequency bands are delivered to basal electrodes. The improvement of speech performance from single-channel to multichannel device demonstrates a clear utilization of place cues in cochlear implant users [1].

In quiet conditions, most cochlear implant users with the latest implant device can achieve a high level of speech performance. However, performance deteriorates significantly in noisy environments [5, 8] even for the best cochlear implant user. The cause of the noise susceptibility of cochlear implant users has been investigated recently. Fu et al. [3] measured the recognition of spectrally degraded vowels and consonants as a function of signal-to-noise ratio in both normal-hearing subjects and cochlear implant users. The results showed that as the spectral information was reduced, speech recognition deteriorated only slightly in quiet conditions, but recognition deteriorated significantly more in noisy conditions. The performance of the best cochlear implant users was similar to that of normal-hearing subjects listening to a similar level of spectral reduction, suggesting that those implant listeners were making optimal use of the spectral cues available. As the spectral resolution was reduced the performance in noise decreased, demonstrating that the limited spectral resolution is a key factor causing the noise susceptibility. However, some of the cochlear implant listeners had poorer performance than processor-matched normal-hearing subjects, suggesting that those implant listeners were not receiving as many spectral channels of information as their number of electrodes, due to unknown factors. One possible additional factor is the loudness mapping function between acoustic amplitude and electric current.

Amplitudes in normal speech can range over 40 to 60 dB. However, implant listeners typically have dynamic ranges of only 6 to 15 dB in electric current, requiring the acoustic range to be compressed into the electric range. Fu and Shannon [2] measured vowel and consonant recognition as a function of the exponent of a power-function nonlinearity in both cochlear implant users and normal-hearing listeners. They found that, for both acoustic and implant listeners, the best performance was obtained when normal loudness was preserved. Performance deteriorated slightly when the amplitude mapping function was either more compressive or more expansive. Thus, instantaneous amplitude nonlinearity only has a minor effect on phoneme recognition in quiet.

The goal of the present study was to understand the effects of nonlinear amplitude mapping on recognition of spectrally degraded speech in noise. The recognition of vowels and consonants was measured in five normal hearing listeners as a function of signal-to-noise ratio, with the exponent of the amplitude-mapping power function as a parameter.

2. METHODS

2.1 Subjects

Five normal-hearing subjects between the ages of 25 and 35 years served as subjects. All subjects had thresholds better than 15 dB HL at audiometric test frequencies from 250 to 8000 Hz and all were native speakers of American English.

2.2 Test materials and procedures

Speech recognition was assessed for medial vowels and consonants. Vowel recognition was measured in a 12-alternative identification paradigm, including 10 monophthongs and 2 diphthongs, presented in a /h/-vowel-/d/ context (heed, hawed, head, who'd, hid, hood, hud, had, heard, hoed, hod, hayed). The

0-7803-5041-3/99 $10.00 © 1999 IEEE

tokens for these closed-set tests were digitized natural productions from 5 male, 5 female, 5 children, drawn from the material collected by Hillenbrand et al. [4]. Consonant recognition was measured in a 16-alternative identification paradigm, for the consonants /b d g p t k l m n f s ʃ v z j θ/, presented in an /a/-consonant-/a/ context. Two repetitions of each of the 16 consonants were produced by three speakers (1 male, 2 female) for a total of 96 tokens (16 consonants * 3 talkers * 2 repeats). All test materials were stored on computer disk and were output via custom software to a 16 bit D/A converter (TDT DD1) at a 16-kHz sampling rate. Speech sounds were presented using a Tucker-Davis-Technologies (TDT) AP2 array processor in a host PC connected via an optical interface.

Each test block included 180 tokens for vowel recognition or 96 tokens for consonant recognition. A stimulus token was randomly chosen from all 180 tokens in vowel recognition and from 96 tokens in consonant recognition and presented to the subject. Following the presentation of each token, the subject responded by pressing one of 12 buttons in the vowel test or one of 16 buttons in the consonant test, each marked with one of the possible responses.

All subjects started with a training session. Two extreme mappings were used as training conditions. Each training session included 8 consecutive test blocks with the same mapping condition and the same speech material. Feedback was provided. The order of training conditions (two mapping conditions and vowel/consonant tests) was randomized across subjects. Subjects started the test sessions after all training conditions were finished. In the test sessions, the order of S/N ratio conditions was randomized. The order of the five mapping conditions, and the order of the vowel and consonant tests, was counterbalanced across subjects. No feedback was provided in test sessions.

2.3 Signal processing

The speech signal was mixed with simplified speech spectrum-shaped noise (constant spectrum level below 800 Hz and 10-dB/octave role-off above 800 Hz). The signal-to-noise ratio (S/N) was defined as the difference, in decibels, between the RMS levels of the whole speech token and the noise. The speech signal was mixed with the noise at S/N levels of 24 dB, 18 dB, 12 dB, 6 dB, 0 dB, -6 dB, -12 dB, for a total of 8 conditions in addition to the original speech.

The spectrally degraded speech stimuli were implemented as follows. The unprocessed speech with the desired S/N level was first pre-emphasized using a first-order Butterworth high-pass filter with a cutoff frequency of 1200 Hz, and then band-pass filtered into 16 frequency bands using eighth-order Butterworth filters. The corner frequencies (3 dB down) of the bands were at 300, 379, 473, 583, 713, 866, 1046, 1259, 1509, 1804, 2152, 2561, 3043, 3612, 4281, 5070, and 6000 Hz. The envelope in each band was extracted by half-wave rectification and low-pass filtering (eighth-order Butterworth) with a 160-Hz cutoff frequency. The envelope was then distorted by a power-law transformation, applied to envelope amplitudes between the maximum envelope value and the noise floor. The exponent of the power function varied from a strongly compressive ($p=0.3$) to a strongly expanded value ($p=3.0$). This distorted envelope of each band was used to modulate a wideband noise, which was then spectrally limited by the same bandpass filter used for that analysis band. The output from all bands were then summed, tokens were equated in rms energy, and presented to the listeners diotically through Sennheiser HDA200 headphones at 70 dBA.

3. RESULTS

Figure 1 shows the mean scores of vowel and consonant recognition as a function of the number of training blocks for the extremely compressed and expanded conditions.

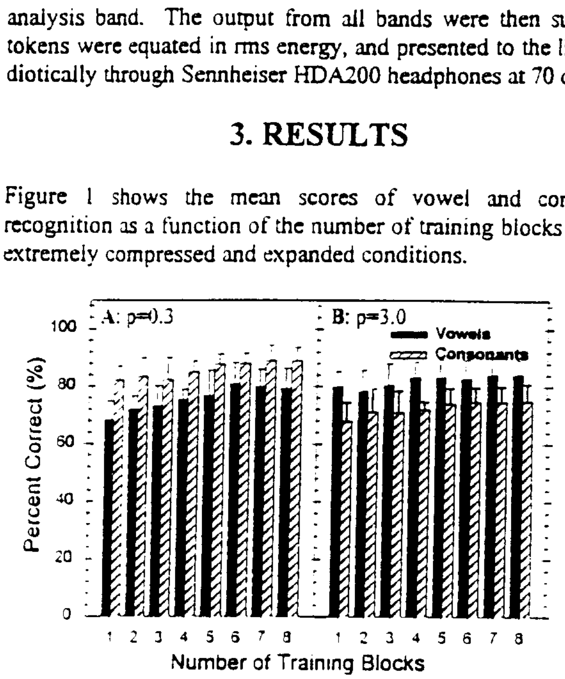


Figure 1. Percent correct of vowel and consonant recognition as a function of the number of training blocks. (A) $p=0.3$; (B) $p=3.0$. Error bars represent +/- one standard deviation.

For the compressed condition ($p=0.3$), the vowel score increased from 68.4% to 80.8%, and consonant scores increased from 82.2% to 88.9% over the eight training sessions, but these increases were not significant [$F(7,32)=2.20$, $p=0.06$ for vowels; $F(7,32)=0.62$, $p=0.74$ for consonants]. However, there was a significant interaction between training and subjects, reflecting a large increase with training for some subjects and no change with training for others [$F(4,35)=8.31$, $p<0.001$ for vowels; $F(4,35)=2.82$, $p=0.04$ for consonants]. For the expanded condition, a 4.4% and 7.0% improvement was observed in vowel and consonant recognition, respectively, but these differences were also not significant [$F(7,32)=1.49$, $p=0.21$ for vowels; $F(7,32)=0.86$, $p=0.55$ for consonants]. Again, there was a significant interaction between subjects and training [$F(4,35)= 21.34$, $p<0.001$ for vowels; $F(4,35)=6.39$, $p<0.001$ for consonants].

Figure 2 shows the mean vowel and consonant recognition scores as a function of the exponent of the power function in quiet and noise condition. In the quiet condition (filled circles), both vowel and consonant scores decreased slightly when either a compressed or expanded mapping was applied. Vowels were relatively more tolerant to expansion while consonants were more tolerant to compression. There was a significant effect of amplitude mapping on recognition of vowels [$F(4,20)=11.49$, $p<0.001$] and consonants [$F(4,20)=23.67$, $p<0.001$]. Post-hoc tests according to Tukey HSD multiple comparisons showed that only the extreme compression ($p=0.3$) significantly reduced the performance in vowel recognition relative to linear mapping ($p=1.0$). Consonant recognition deteriorated significantly in all mapping conditions except the moderate compression ($p=0.5$). In noise conditions, amplitude mapping had a significant impact on vowel and consonant recognition at all signal-to-noise levels.

Post-hoc Tukey HSD tests showed no significant performance drop in conditions with expansive mapping relative to linear mapping. Indeed, the extreme expansion (p=3.0) significantly improved the vowel recognition scores at high noise levels (-6 dB SNR). Post-hoc Tukey HSD tests also showed a significant performance drop in all conditions at all noise levels with compressive mappings relative to those with linear mappings.

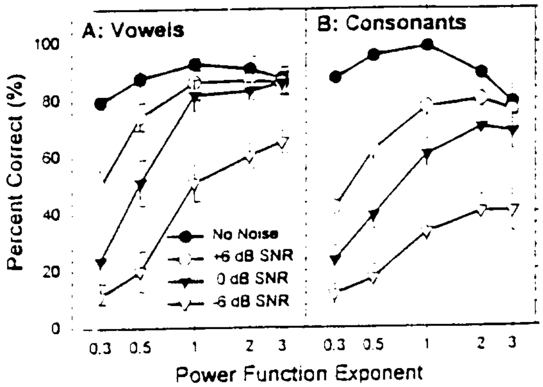


Figure 2. Recognition scores of vowels and consonants as a function of the exponent of the power function in quiet and noise condition. (A) Vowels; (B) Consonants. Error bars represent +/- one standard deviation.

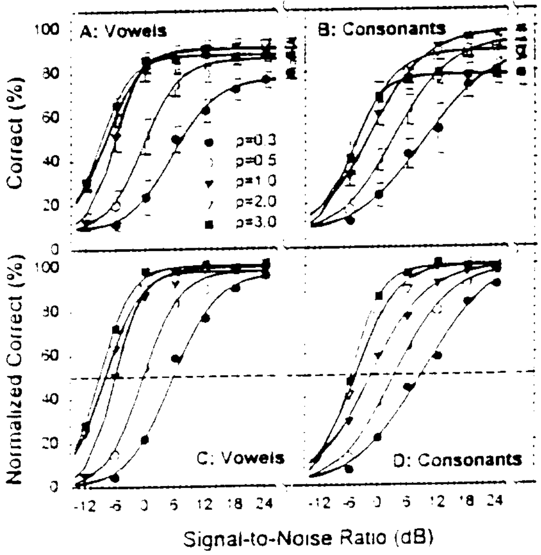


Figure 3. Recognition scores of vowels and consonants as a function of signal-to-noise ratio. (A) Vowel scores; (B) Consonant scores; (C) Normalized vowel scores; (D) Normalized consonant scores. The solid lines represent the fitting curve based on the Equation 1 and experimental data. The dashed lines represent 50% levels. Error bars represent +/- one standard deviation.

Figures 3A and 3B show the mean scores of vowel and consonant recognition, respectively, as a function of S/N ratio with different amplitude mappings. Both vowel and consonant scores gradually decreased as signal-to-noise (S/N) ratio decreased for all mapping conditions. Figures 3C and 3D show the normalized performance on vowels and consonants, respectively, as a function of S/N ratio, relative to the performance in quiet. The dashed lines in Figures 3C and 3D indicate 50% of the normalized score after correction for chance. The S/N level that produced this 50% level of performance was defined as the phoneme recognition threshold (PRT).

The data of Figure 3 were fit by a simple sigmoidal model.

$$\%C = P_0 + (Q - P_0)/(1+\exp(-\beta(x-PRT))) \qquad (1)$$

where Q is the percent correct in quiet, PRT is the phoneme recognition threshold in dB, x is the S/N ratio in dB, P_0 is the chance level (6.25% for consonants, 8.33% for vowels), and β is related to the slope of the function at PRT. Figure 4 shows the PRTs and slopes as a function of the power function exponents. The fits of this function to the data were uniformly excellent, with all r^2 values better than 0.99. The PRT for both vowels and consonants improved significantly as the mapping function changed from a compressive mapping to an expanded mapping [F(4,40)=190.14, p<0.001]. The slopes of the vowel and consonant functions at PRT also changed significantly as a function of the mapping exponent [F(4,40)=8.03, p<0.001].

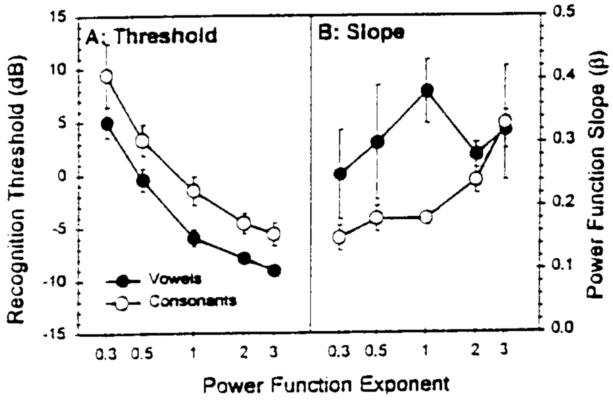


Figure 4. Phoneme recognition threshold and the slope of vowel and consonant recognition as a function of the power function exponents. (A) Phoneme recognition threshold; (B) The slope (β) of the power function. Error bars represent +/- one standard deviation.

4. DISCUSSION

In normal hearing the signal of interest (speech) and the interfering noise are processed through the same sensory channels in the normal cochlea. However, in patients with hearing impairment or in deaf patients with cochlear implants the signal stream must be pre-processed to remediate the impaired sensory processing. Processing strategies need to accommodate not only ideal conditions of listening in quiet, but also real-world conditions of listening in high noise levels. Some modern digital hearing aids and all cochlear implants use instantaneous nonlinear amplitude compression to restore normal loudness sensations to the listener with impaired sensory processing. The present study measured phoneme recognition in quiet and in noise as a function of the nonlinear amplitude mapping. Spectral resolution was reduced to 16 bands to simulate the reduced spectral resolution in an impaired cochlea. Although the results are most directly applicable to signal processing for hearing-impaired listeners, they also have interesting implications for normal hearing in noisy conditions.

The present results replicate the finding [2, 6] that amplitude nonlinearity in quiet has only a minor effect on phoneme recognition. However, as the S/N level decreases, the effect of amplitude nonlinearity becomes dramatic and asymmetric: expansive mappings are only mildly effected by noise, while compressive mappings are strongly effected. One implication of this result is that expansive mappings may be better overall for mixed quiet and noisy conditions. The expansive exponents may be slightly poorer in quiet conditions, but would still allow reasonably good speech recognition in noise. In contrast, compressive mappings would allow a similar level of speech recognition in quiet, but would be considerably worse in noise. An interesting implication is that a processor with an expansive nonlinearity may improve speech recognition in noise compared to no processing even for normal-hearing subjects.

The results in the present experiment indicated that the improvement by learning was subject-dependent as well as stimulus-dependent. The improvement in the present study was much less than that reported by Licklider and Pollack [6], which may be simply due to the difference in test materials. The improvement of consonant recognition was similar for either compressed or expanded speech. However, more improvement in vowel recognition was observed for the compressed speech than the expanded speech. Some subjects improved more than 20% after 8 training blocks, while other subjects showed no improvement with the same training. The variation of training effects across subjects was unexpectedly large. Possible reasons may include the motivation of subjects or training procedure.

Although amplitude distortion has only a small affect on speech intelligibility in quiet [3, 6], Thomas and Niederjohn [9,10] found that amplitude-compressed speech was recognized at a much higher level than uncompressed speech at high noise levels. This result appears to be contradictory to the results in the present study, which showed a devastating effect of amplitude compression on speech intelligibility in noise. In Thomas and Niederjohn's experiments amplitude compression was applied to the noise-free speech to which uncompressed noise was then added. These earlier methods are applicable where the noise-free speech is available for processing, prior to the introduction of noise. However, their method is not appropriate for most listening situations in everyday life where the speech and noise are added together before the processing can be applied.

The present results also show an interesting difference between vowel and consonant recognition. In the quiet condition, the influence of amplitude compression on vowel and consonant recognition was similar. However, consonant recognition deteriorated much faster than vowel recognition for expanded speech. Further analysis showed that performance on the manner cues suffered most [7]. Amplitude mapping had a similar impact on the recognition pattern of the PRTs for vowels and consonants although the slope of the sigmoidal functions was different.

The data in the present study showed that the PRT was highly dependent on amplitude mapping. Slightly expansive mapping may be better overall in combined quiet and noisy listening conditions. Compressive mapping functions may be satisfactory in quiet, but result in a large decrease in performance in noise. This suggests that at least part of the high variability in cochlear implant users may be due to non-optimal amplitude mapping. Implant listeners who have an amplitude mapping function that is too compressive would be at a disadvantage in noise compared to implant listeners with expansive loudness mappings. The asymmetry of these results suggests that a slightly expansive mapping might be the best choice for overall listening conditions.

5. SUMMARY AND CONCLUSIONS

Nonlinear amplitude mapping produced only a mild decrement in speech recognition in quiet, but could produce a large decrement in noise. Expansive nonlinear mapping provides better overall performance in noise than linear or compressive mapping.

6. ACKNOWLEDGMENTS

We wish to thank Professor James Hillenbrand for allowing us to use the multi-talker vowel test materials. The assistance of Xiaosong Wang in collecting data was also gratefully acknowledged. The research was supported by NIDCD.

7. REFERENCES

[1] Fishman, K., Shannon, R.V., and Slattery, W.H. (1997). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor." J. Speech Hear. Res. 40, 1201-1215.

[2] Fu, Q.-J., and Shannon, R.V. (1998). "Effects of amplitude nonlinearity on speech recognition by cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am., 104(4).

[3] Fu, Q.-J., Shannon, R.V., Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am., submitted.

[4] Hillenbrand J., Getty, L.A., Clark, M.J., and Wheeler K. (1995). "Acoustic characteristics of American English vowels." J. Acoust. Soc. Am. 97, 3099-3111.

[5] Hochberg, I., Boothroyd, A., Weiss, M., Hellman, S. (1992). "Effects of noise and noise suppression on speech perception by cochlear implant users." Ear and Hearing 13, 263-271.

[6] Licklider, J.C.R. and I. Pollack (1948). "effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech," J. Acoust. Soc. Am. 20, 42-51.

[7] Miller, G. and Nicely, P. (1955). "An analysis of perceptual confusions among some English consonants." J. Acoust. Soc. Am. 27, 338-352.

[8] Müller-Deiler, J., Schmidt, B.J., and Rudert, H. (1995). "Effects of noise on speech discrimination in cochlear implant patients." Ann. Otol. Rhinol. Laryngol. 166, 303-306.

[9] Niederjohn, R.J. and J.H. Grotelueschen. "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression." IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 24, pp. 277-282, Aug. 1976.

[10] Thomas, I.B. and R.J. Niederjohn. "The intelligibility of filtered-clipped speech in noise." Journal of the Audio Engineering Society, Vol. 18, pp. 299-303, June, 1970.

Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping

Qian-Jie Fu and Robert V. Shannon

Department of Auditory Implants and Perception, House Ear Institute,
2100 West Third Street, Los Angeles, CA 90057

qfu@hei.org *shannon@hei.org*

Abstract: The present study measured phoneme recognition as a function of signal-to-noise levels when different nonlinear loudness mapping functions were implemented in three cochlear implant users using a 4-channel CIS speech processing strategy. Results showed that phoneme recognition scores in quiet varied only slightly when different amplitude mappings were applied, from highly compressive to weakly compressive. As the level of background noise was increased, recognition scores decreased more rapidly for the strongly compressive mapping than for the weakly compressive mapping. Results indicate that, although a strongly compressive mapping between acoustic and electric amplitude produced slightly better performance in quiet, a less compressive mapping may be beneficial in noisy listening conditions for cochlear implant listeners.

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts

1. Introduction

In quiet laboratory testing conditions, many cochlear implant users with the latest implant devices can achieve high levels of open-set sentence recognition. However, performance deteriorates significantly in noisy listening conditions (Hochberg et al., 1992; Müller-Deiler et al., 1995) even for the best cochlear implant users. Several explanations for the noise susceptibility of cochlear implant listeners have been proposed recently. One of the most obvious factors is the limited spectral resolution in cochlear implants. In a recent study, Fu et al. (1998) measured phoneme recognition in five normal-hearing listeners as a function of the number of spectral channels. Results showed that as the spectral information was reduced, speech recognition deteriorated only slightly in quiet, but recognition deteriorated significantly in noise. Phoneme recognition performance of the best cochlear implant users was similar to that of normal-hearing subjects listening to a similar level of spectral reduction. A similar result was reported by Dorman and colleagues (1998a, 1998b). These results indicated that the limited spectral resolution in cochlear implant listeners is a key factor causing noise susceptibility. However, one interesting observation from both Fu et al. and Dorman et al. studies is that some implant listeners had performance comparable to normal-hearing listeners in quiet but had significantly poorer performance in noise, even when using a similar speech processor. One factor that may have contributed to this difference is the loudness mapping function between acoustic amplitude and electric current.

Instantaneous amplitudes in normal speech range over as much as 60 dB. However, implant listeners typically have dynamic ranges of only 6 to 15 dB in electric current, requiring the acoustic range to be compressed into the electric range. Zeng and colleagues (1994, 1999) determined that loudness in electrical stimulation could be represented by an exponential function of current. This function implied that a highly compressive function is necessary to map acoustic amplitudes to electric amplitudes to preserve loudness. Fu and Shannon (1998) investigated the effect of nonlinear amplitude mapping on vowel and consonant recognition in both cochlear implant users and normal-hearing listeners. They found that, for both acoustic and implant listeners, the best performance was obtained when normal loudness was preserved which, for electrical stimulation, was obtained when a compressive power-law mapping (p=0.22) was applied. A traditional power-law, cross-modality model indicated that this mapping best restored the loudness growth in cochlear implant users. Performance deteriorated only slightly in both acoustic and implant listeners when the amplitude mapping function was either more compressive or more expansive. Thus, instantaneous amplitude nonlinearity only has a minor effect on phoneme recognition in quiet.

Fu and Shannon (1999) investigated the effects of nonlinear amplitude mapping on the recognition of spectrally degraded speech in noise by normal-hearing subjects. They measured vowel and consonant recognition in five normal-hearing listeners as a function of signal-to-noise ratio, with the exponent of the amplitude-mapping power function as a parameter. The results showed that nonlinear amplitude mapping produced only a mild decrement in speech recognition in quiet, but could produce a large decrement in noise. Expansive nonlinear mapping provided better overall performance than linear or compressive mapping in low SNR conditions.

The goal of the present study is to investigate the effect of nonlinear amplitude mapping on phoneme recognition in cochlear implant users.

2. Methods

2.1 Subjects

Cochlear implant subjects were three post-lingually deafened adults using the Nucleus-22 device. All had at least four years experience utilizing the SPEAK speech processing strategy and all were native speakers of American English. All implant subjects had 20 active electrodes available for use. Two subjects (N4 and N7) used frequency allocation table 9 (150-10,823 Hz) in their clinical implant processor and one subject (N3) used frequency allocation table 7 (120 Hz - 8,658 Hz). All implant participants had extensive experience in speech recognition experiments. Table 1 contains relevant information for the three subjects, including their most recent scores on the CUNY sentence test and on a multi-talker, 12-vowel recognition test with their 20-electrode SPEAK processor (McDermott et al., 1992).

Table 1. Subject information three five Nucleus-22 cochlear implant listeners who participated in the present study.

Subject	Age	Gender	Cause of Deafness	Duration of use	Freq. Table	Score (CUNY)	Vowel Score
N3	56	M	Trauma	7 years	7	96.2%	69.5%
N4	40	M	Trauma	5 years	9	100.0%	81.1%
N7	55	M	Unknown	5 years	9	100.0%	64.5%

2.2 Test materials and procedures

Speech recognition was assessed for medial vowels and consonants. Vowel recognition was measured in a 12-alternative identification paradigm, including 10 monophthongs and 2 diphthongs, presented in a /h/-vowel-/d/ context (heed, hawed, head, who'd, hid, hood, hud, had, heard, hoed, hod, hayed). The tokens for these closed-set tests were digitized natural productions from 5 males, 5 females, and 5 children, drawn from the material collected by Hillenbrand et al. (1995). Consonant recognition was measured in a 16-alternative identification paradigm, for the consonants /b d g p t k l m n f s ʃ v z j θ/, presented in an /a/-consonant-/a/ context. Two repetitions of each of the 16 consonants were produced by three speakers (1 male, 2 female) for a total of 96 tokens (16 consonants * 3 talkers * 2 repeats).

Each test block included 180 tokens for vowel recognition or 96 tokens for consonant recognition. A stimulus token was randomly chosen from all 180 tokens in vowel recognition and from 96 tokens in consonant recognition and presented to the subject. Following the presentation of each token, the subject responded by pressing one of 12 buttons in the vowel test or one of 16 buttons in the consonant test, each marked with one of the possible responses.

All signals were presented at comfortable audible levels based on the CIS speech processing strategy through a custom implant interface system (Shannon et al., 1990). Subjects were well familiarized with the test materials and the test procedure from prior experiments. All subjects started with a training session. Speech sounds without noise were used as training conditions. Each training session included 8 consecutive test blocks with the same mapping condition and the same speech material. Feedback was provided. Subjects started the test sessions after either 8 consecutive runs or the performance was stabilized in consecutive three runs. In the test sessions, the order of S/N ratio conditions was randomized. The order of the five mapping conditions, and the order of the vowel and consonant tests, were counterbalanced across subjects. No feedback was provided in test sessions.

2.3 Signal processing

The speech signal was mixed with simplified speech spectrum-shaped noise (constant spectrum level below 800 Hz and 10-dB/octave role-off above 800 Hz). The signal-to-noise ratio (S/N) was defined as the difference, in decibels, between the RMS levels of the whole speech token and the noise.

The 4-channel Continuous Interleaved Sampling (CIS) processor (Wilson et al., 1991) was implemented as follows. The signal was first pre-emphasized using a first-order Butterworth high-pass filter with a cutoff frequency of 1200 Hz, and then band-pass filtered into four broad frequency bands using eighth-order Butterworth filters. The five corner frequencies of the four bands were at 300 Hz, 713 Hz, 1509 Hz, 3043 Hz, and 6000 Hz. The envelope of the signal in each band was extracted by half-wave rectification and low-pass filtering (8th order Butterworth) with a 160 Hz cutoff frequency. The amplitude histogram in each band was computed for the test materials presented at 70 dB SPL. The maximum amplitude used (A_{max}) was set to the 99th %-ile of all amplitude levels in all channels and the minimal amplitude (A_{min}) was set to the noise floor in the absence of sound input in all channels. The current level (E) of electric stimulation in the ith band was set to the acoustic envelope value (A) raised to a power (Fu and Shannon, 1998). The exponent of the power function was systematically changed from 0.05 to 0.8. This transformed amplitude was used to modulate the amplitude of a continuous, 500 pulse/sec, biphasic pulse train with a 100 μs/phase pulse duration. The stimulus order of the 4 channels was 1-3-2-4 for electrode pairs (16,22), (11,17), (6,12), and (1,7), respectively.

3. Results

Figure 1A and 1B show the mean and individual scores of vowel and consonant recognition as a function of the power function exponents in the quiet condition. The dotted lines show the individual scores from three listeners and the solid line shows the mean scores from these three subjects. The mean vowel scores were consistently recognized at about 50% correct when the exponent of the power function was increased from 0.05 to 0.4, and slightly dropped to 41.7% as the exponent of the power function further increased to 0.8. Similarly, the mean consonant scores changed slightly from 70% when the value of the exponent, p, of the amplitude mapping function was 0.05, to 73% when p was 0.2, and dropped to 46% when the exponent was increased to 0.8.

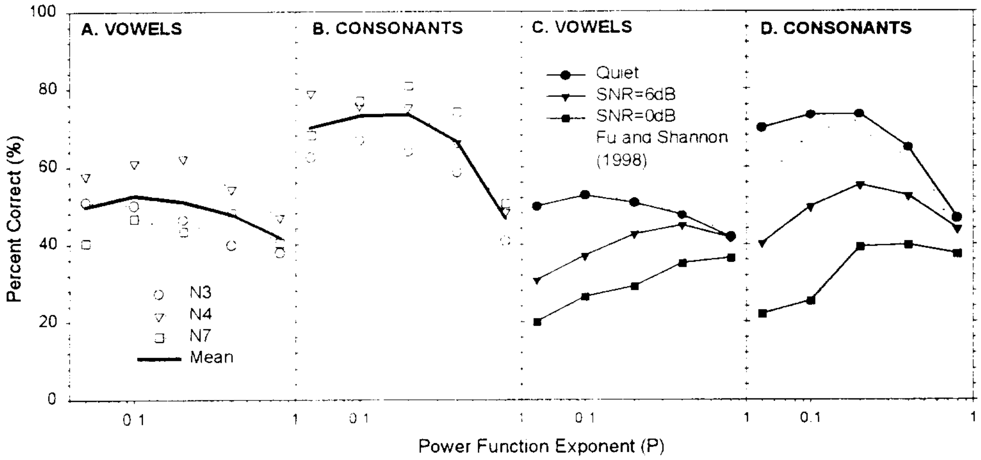


Fig. 1. Recognition scores of vowels and consonants as a function of the power function exponents in quiet and in noise. (A) Vowels (in quiet condition), (B) Consonants (in quiet condition); (C) Vowels (mean scores); (D) Consonants (mean scores).

Figure 1C and 1D show the mean vowel and consonant recognition scores as a function of the exponent of the power function in quiet and in noise. For weak compression (p=0.8), only a slight drop of speech performance was observed at both +6 dB SNR and 0 dB. However, a much larger reduction in performance was observed as the noise level increased for the strongly compressive conditions. When the exponent p was 0.05, a 20% reduction occurred in vowel recognition and 30% reduction in consonant recognition were observed at going from quiet conditions to +6 dB SNR.

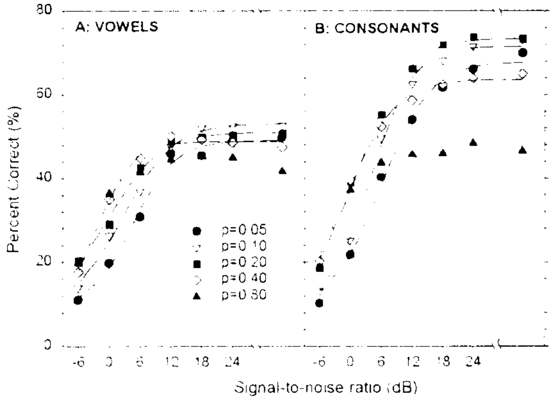


Fig. 2. Recognition scores of vowels and consonants as a function of signal-to-noise ratio. (A) Vowels; (B) Consonants. The lines represent the fitting curve based on a sigmoidal model.

Figures 2A and 2B show the mean scores of vowel and consonant recognition, respectively, as a function of S/N ratio with different amplitude mappings. Both vowel and consonant scores gradually increased as signal-to-noise (S/N) ratio increased for all mapping conditions. The phoneme recognition threshold (PRT) was defined as the S/N level that produced 50% of the performance level in quiet. The lines represent the best fit of a simple sigmoidal model (Fu et al., 1998; Fu and Shannon, 1999) with three parameters: PRT, the slope of the function at PRT (β), and the performance level in quiet.

Figure 3 shows the PRT as a function of the power function exponent. The PRT for both vowels and consonants improved significantly as the mapping function changed from a strong compression (p=0.05) to a weak compression (p=0.8).

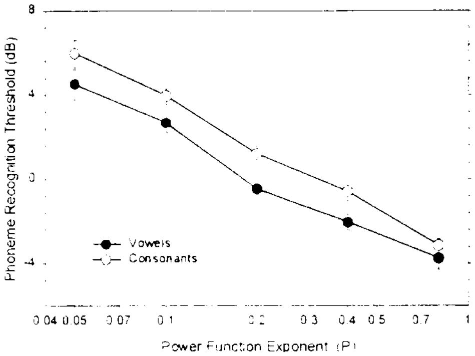


Fig. 3. Phoneme recognition threshold of vowel and consonant recognition as a function of the power function exponents. Error bars represent + - one standard deviation.

4. Discussion and conclusions

The present results demonstrate that nonlinear acoustic-to-electric amplitude mappings have only a minor effect on phoneme recognition in quiet, consistent with the previous finding in normal-hearing listeners (Fu and Shannon, 1998). However, as the S/N level decreases, the effect of nonlinear amplitude mapping becomes dramatic and asymmetric: performance with weakly compressive mappings declines mildly in noise, but performance declines dramatically in noise with a strongly compressive amplitude mapping.

One interesting point of comparison in the present results is the effect of stimulation mode. The present results, obtained with BP+5 stimulation mode, are quite similar to previous results (Fu and Shannon, 1998) in the same implant listeners with BP+1 stimulation mode (Figure 1C and 1D, dotted lines). Stimulation in BP+5 mode should produce a broader region of stimulated nerves than BP+1 mode. The comparison in Figure 1C and 1D shows that vowel and consonant recognition in quiet with BP+5 mode is even less affected by changes in the degree of amplitude mapping nonlinearity than with BP+1 mode. When compared to the condition with the highest performance, the strongly compressive mapping (p=0.05) resulted in a 16% reduction of vowel and consonant recognition with BP+1 stimulation mode, but only a 3% reduction with BP+5 stimulation mode. One possible explanation for this difference might be the different loudness growth functions for the two stimulation modes. Chatterjee (1999) found that the loudness growth function was steeper for electrode pairs with broad stimulation mode than narrow stimulation mode. The power law exponent that produces the "optimal" loudness mapping should be slightly more compressive for broad stimulation than for narrow stimulation. If this were the case we would expect the entire phoneme recognition function to shift (left) to more compressive exponent values for the wider stimulation mode. Contrary to this prediction, phoneme recognition performance was similar for the two stimulation modes for weak compression (p=0.8). Another possibility is that wider stimulation modes are less affected by amplitude mapping and so the entire phoneme recognition function would be flatter. Again the data match this prediction at highly compressive exponents but not at weakly compressive exponents. Since the pattern of results is not consistent with either of these predictions, the explanation for the difference in performance for different stimulation modes is still not clear.

The present results also show an interesting similarity between vowel and consonant recognition. In quiet, neither vowel nor consonant recognition was strongly affected by amplitude compression, although consonant recognition did deteriorate more than vowel recognition for the most linear mapping (p=0.8). Further analysis showed that performance on manner cues (Miller and Nicely, 1955) suffered most in this condition. The amplitude mapping exponent had a similar effect on the PRT for vowels and consonants (Figure 3).

The data in the present study showed that the PRT was highly dependent on amplitude mapping. This suggests that at least part of the large variability in performance across cochlear implant users may be due to non-optimal amplitude mapping. Implant listeners who have an amplitude mapping function that is too compressive would be at a disadvantage in noise compared to implant listeners with less compressive mappings. One implication of the results is that less compressive mappings may be better overall for mixed quiet and noisy conditions. Amplitude mappings that are more linear than the optimal loudness mapping may be slightly poorer in quiet conditions, but would still allow reasonably good speech recognition in noise. In contrast, strongly compressive mappings would allow a similar level of speech recognition in quiet, but would be considerably worse in noise.

Acknowledgments

We wish to thank Prof. James Hillenbrand for allowing us to use the multi-talker vowel test materials. The research was supported by grants R03-DC-03861, R01-DC-01526, and contract N01-DC-92100 from NIDCD.

References and links:

Chatterjee, M. (**1999**). "Effects of stimulation mode on threshold and loudness growth in multielectrode cochlear implants," J. Acoust. Soc. Am. **105**, 850-860.

Dorman, W.F., Loizou, P.C., and Fitzke, J. (**1998a**). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," Ear and Hearing **19**, 481-484.

Dorman, W.F., Loizou, P.C., Fitzke, J., and Tu, Z. (**1998b**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583-3585.

Fu, Q.-J., and Shannon, R.V. (**1998**). "Effects of amplitude nonlinearity on speech recognition by cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **104**, 2571-2577.

Fu, Q.-J., Shannon, R.V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586-3596.

Fu, Q.-J. and Shannon, R.V. (**1999**). "Recognition of spectrally degraded speech in noise with nonlinear amplitude mapping," Proceedings of 1999 IEEE International Conference of Acoustics, Speech, and Signal Processing, Vol. 1, 369-372.

Hillenbrand J., Getty, L.A., Clark, M.J., and Wheeler K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099-3111.

Hochberg, I., Boothroyd, A., Weiss, M., Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear and Hearing **13**, 263-271.

McDermott, H.J., McKay, C.M., and Vandali, A.E. (**1992**). A new portable sound processor for the University of Melbourne/Nucleus Limited Multichannel cochlear implant, J. Acoust. Soc. Amer., **91**, 3367-3371.

Miller, G. and Nicely, P. (**1955**) "An analysis of perceptual confusions among some English consonants." J. Acoust. Soc. Am. **27**, 338-352.

Müller-Deiler, J., Schmidt, B.J., and Rudert, H. (**1995**). "Effects of noise on speech discrimination in cochlear implant patients," Ann. Otol. Rhinol. Laryngol. **166**, 303-306.

Shannon, R.V., Adams, D.D., Ferrel, R.L., Palumbo, R.L., and Grantgenett, M. (**1990**). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant." J. Acoust. Soc. Am. **87**, 905-907.

Wilson, B. S., Finley, C. C. Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (**1991**). New levels of speech recognition with cochlear implants, Nature, **352**, 236-238.

Zeng, F.-G. and Shannon, R.V. (**1994**). "Loudness coding mechanisms inferred from electric stimulation of the human auditory system", Science, **264**, 564-566

Zeng, F.-G. and Galvin, J. (**1999**). Amplitude mapping and phoneme recognition in cochlear implant listeners, Ear & Hearing, **20(1)**, 60-74.

Effect of stimulation rate on phoneme recognition in cochlear implants

Qian-Jie Fu
Robert V. Shannon

Department of Auditory Implants and Perception
House Ear Institute, 2100 West Third Street
Los Angeles, CA 90057

Draft: December 15, 1998

Submitted to Journal of Acoustical Society of America

Received:____

Send Correspondence to:
Qian-Jie Fu, Ph.D.
Department of Auditory Implants and Perception
House Ear Institute, 2100 West Third Street
Los Angeles, CA 90057
Phone: (213) 273-8036
FAX: (213) 413-0950
E-mail: qfu@hei.org

ABSTRACT

This study investigated the effect of pulsatile stimulation rate on medial vowel and consonant recognition in cochlear implant listeners. Experiment 1 measured phoneme recognition as a function of stimulation rate in three Nucleus-22 cochlear implant listeners using a custom 4-channel CIS speech processing strategy. Results showed that all stimulation rates of 150 pulses/sec/electrode or higher produced equally good performance, while stimulation rates lower than 150 pulses/sec/electrode produced significantly poorer performance. Experiment 2 measured phoneme recognition by implant listeners and normal-hearing listeners as a function of the low-pass cutoff frequency for envelope information. Results from both acoustic and electric hearing showed no significant difference in performance for all cutoff frequencies higher than 20 Hz. Both vowel and consonant scores dropped significantly when the cutoff frequency was reduced from 20 Hz to 2 Hz. The results of these two experiments suggest that temporal envelope information can be conveyed by relatively low stimulation rates. The pattern of results for both electrical and acoustic hearing is consistent with a simple model of temporal integration with an equivalent rectangular duration (ERD) of the temporal integrator of about 7 ms.

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts.

INTRODUCTION

The continuous interleaved sampling (CIS) stimulation strategy is one of the most successful speech processing strategies for electrical stimulation of the auditory nerve. For speech processors with the CIS strategy, the spectral representation of speech sounds is encoded coarsely by the spatial location of the stimulated electrodes and good temporal resolution is achieved by high-rate stimulation with short biphasic pulses interleaved in time across electrodes (Wilson et al, 1991).

The two major parameters of the CIS strategy are the number of stimulation channels and the rate of stimulation in each channel. It is widely assumed that increasing the number of channels and/or stimulation rate would improve speech perception. A greater number of channels would provide better spectral resolution of speech sounds, while a higher rate of stimulation would improve the temporal sampling of the speech signal and possibly increase the "stochastic" response properties of the activated neurons (Wilson et al., 1997, 1998). Because the CIS strategy is a non-simultaneous strategy that allows only one channel to be stimulated at any time to avoid electrical field interaction across electrodes, there is an inherent tradeoff between the number of channels and the stimulation rate on each channel.

Brill et al. (1997) investigated the tradeoff between the number of channels and the rate of stimulation in three cochlear implant listeners using the Med-El COMBI 40+ device. Two experiments were conducted in their study. In one experiment, the stimulation rate per channel increased as the number of channels decreased to keep the overall rate constant at 18181 pps. In the second experiment, the stimulation rate per channel was kept constant at 1515 pps regardless of the number of channels. They found little change in performance as the number of channels increased from 4 to 12 in both experiments for one subject. However, trading channels for higher stimulation rate did improve performance for one of the three subjects, who seemed to have maximum performance at 6 and 8 channels. They suggested that a good fitting strategy for a cochlear implant processor might be to select the six or eight "best" channels and let them operate at a high stimulation rate rather than select a large number of channels which would necessitate a lower stimulation rate.

Fishman et al. (1997) conducted another systematic study of the effect of the number of channels on speech performance. They measured speech recognition of eleven subjects with the Nucleus 22 implant and the SPEAK processor as a function of the number of electrodes. In their method, the stimulation rate per electrode increased as the number of channels decreased. Their results showed no significant difference in performance between 7-, 10- and 20-electrode processors, which represent maximal stimulation rates of approximately 750, 500, and 250 pps per electrode, respectively. It is possible, but unlikely, that an improvement due to the increase in the number of electrodes was exactly offset by the decrement due to lower the stimulation rate.

Several studies have been conducted to investigate the effect of stimulation rate on speech performance with a fixed number of channels. Lawson et al. (1996) measured consonant recognition in five cochlear implant listeners using a six-channel CIS strategy. Two different stimulation rates, 250 pulses/sec (pps) and 2525 pps-per-channel, were used in their study. Four out of five subjects did show some evidence of performance improvement when the stimulation rates were increased from 250 to 2525 pps-per-channel, but magnitude of the improvement was small.

Recently, Vandali et al. (1998) evaluated the effect of varying the stimulation rate on speech perception in five postlinguistically deaf adult users of the Nucleus 24M Cochlear Implant

System with the SPEAK processing strategy. In their study, three different rates of electrical stimulation (250, 807, and 1615 pulses per second per channel) were employed on all 20 active electrodes. They found no statistically significant difference in performance between the low and medium stimulation rates. However, significantly poorer results were observed for the high-rate condition for some tests with some individuals.

While these early results showed relatively little effect of stimulation rate on speech recognition, it seems clear that higher pulse rates can have a profound effect on the firing patterns of individual neurons. Wilson et al. (1997, 1998) demonstrated from intra-cochlear compound action potential recordings that high stimulation rates can disrupt the extreme synchrony observed with low-rate electrical stimulation, allowing more normal "stochastic" firing patterns. They suggested that when pulses are presented at high rates, low levels of neural membrane noise at nodes of Ranvier might interact with the pulses to produce stochastic independence among neurons. Slight variations in neural threshold due to membrane noise may introduce a "jitter" in firing times across neurons for rapidly presented pulses. After the onset of a pulse train, this jitter would be expected to increase with time, due to the initial differences in neural discharge histories, combined with the probabilistic recovery time of each neuron. After a short period of time, these differences in discharge histories may produce a high level of stochastic independence among neurons. Thus, it is clear that high rates of electrical stimulation can produce more stochastic patterns of temporal discharges that lower rates, but the implication of this for speech recognition is not clear.

These previous studies present a mixed picture of the effect of stimulation rate upon speech recognition in implant listeners. The present study systematically investigates the effect of stimulation rate, especially for very low stimulation rates, on phoneme recognition in cochlear implant listeners. The allocation of frequencies to electrodes and the electrode locations are held constant in these processors; only the rate of stimulation is varied. Vowel and consonant scores are measured as a function of stimulation rate.

I. EXPERIMENT 1: EFFECT OF STIMULATION RATE

This experiment measured vowel and consonant recognition as a function of stimulation rate in cochlear implant listeners using speech processors fitted with a custom four-channel CIS speech processing strategy. Seven stimulation rate conditions were tested in the present study: 50, 100, 150, 200, 300, 400, and 500 pps-per-channel.

A. METHODS

1. Subjects

Three post-lingually deafened adults using the Nucleus-22 cochlear implant device participated in this study. All were native speakers of American English and had at least four years experience utilizing the SPEAK speech processing strategy. All implant subjects had 20 active electrodes available for use. Two subjects (N4 and N7) used frequency allocation table 9 (150-10,823 Hz) in their clinical implant processor and one subject (N3) used frequency allocation table 7 (120 – 8,658 Hz). Based on their sentence and word recognition scores, subjects N4 and N7 were considered excellent implant users, and subject N3 was considered an average user. No poor users were chosen in this study to avoid floor effects. All implant participants had extensive experience in speech recognition experiments. Table 1 contains relevant information for the three subjects, including their most recent scores on the CUNY sentence test and multi-talker vowel test, presented without lip-reading in the sound field through their normal 20-electrode SPEAK processor.

--------------------Table 1 about here-----------------

2. Test Materials and Procedures

Speech performance was assessed using two measures: vowel and consonant identification. Vowel recognition was measured in a 12-alternative identification paradigm, including 10 monophthongs (/i ɪ ɛ æ u ʊ ɑ ʌ ɔ ɝ/) and 2 diphthongs (/o e/), presented in a /h/vowel/d/ context. The tokens for these closed-set tests were digitized natural productions from 5 men, 5 women, 3 boys, and 2 girls, drawn from the materials collected by Hillenbrand et al. (1995). Consonant recognition was measured in a 16-alternative identification paradigm, for the consonants / b d g p t k l m n f s ʃ v z θ dʒ /, presented in an /a/consonant/a/ context. Two repetitions of each of the 16 consonants were produced by three speakers (1 male, 2 female) for a total of 96 tokens (16 consonants * 3 talkers * 2 repeats).

All signals were presented at levels that were adjusted to be comfortably loud by each individual subject in each listening session. Subjects were instructed to maintain the same loudness scale across all test conditions. All 14 conditions (2 tests * 7 rates) were pseudo-randomized within subjects.

3. Signal Processing

The 4-channel CIS processor was implemented through the custom interface (Shannon et al., 1990), bypassing the subject's Spectra-22 speech processor. The signal was first pre-emphasized using a first-order Butterworth (6 dB/octave) high-pass filter with a cutoff frequency of 1200 Hz, and then band-pass filtered into four broad frequency bands using eighth-order Butterworth filters (96dB/octave). The corner frequencies of the bands were at 300 Hz, 713 Hz, 1509 Hz, 3043 Hz, and 6000 Hz. The envelope of the signal in each band was extracted by half-wave rectification and low-pass filtering (eighth-order Butterworth: 48 dB/octave). To avoid any aliasing effects, the cutoff frequency of the low-pass envelope filters was set to 40% of the rate of stimulation. The acoustic amplitude (40-dB range) was transformed into electric amplitude by a power-law function with an exponent of 0.2 ($E = A^{0.2}$; Fu and Shannon, 1998) between each subject's threshold (T-level) and upper level of loudness (C-level). This transformed amplitude was then used to modulate the amplitude of a continuous biphasic pulse train with a 100 µs/phase pulse duration, and delivered to four electrode pairs interleaved in time: (18,22), (13,17), (8,12), and (3,7). The rate of the biphasic pulse train was varied from 50 pps/channel to 500 pps/channel, for a total of 7 stimulation rate conditions.

B. RESULTS AND DISCUSSION

Figure 1 shows the vowel and consonant recognition scores as a function of stimulation rate. Panel A shows the individual and mean vowel recognition scores and Panel B shows the individual and mean consonant recognition scores. When the stimulation rate was 50 pps/channel, 41% of vowels and 46% of consonants were correctly recognized. When the stimulation rate was increased from 50 pps to 150 pps/channel, mean vowel scores increased significantly to 53% correct and mean consonant scores increased significantly to 71% correct. No significant improvement of vowel or consonant recognition was observed when the stimulation rate was further increased. For vowel recognition, ANOVA tests showed a significant effect of stimulation rate [F(6,21)=22.20, p<0.001], of subjects [F(2,21)=66.57, p<0.001], and significant interaction between stimulation rate and subjects [F(6,21)=3.20, p=0.010]. For consonant recognition, statistical tests also showed a significant effect of stimulation rate [F(6,21)=103.72, p<0.001], of subjects [F(2,21)=55.65, p<0.001], and significant interaction between stimulation rate and subjects [F(6,21)=7.742, p<0.001]. Post-hoc Tukey HSD tests indicate that both vowel and consonant scores were significantly lower for speech processors with stimulation rates lower than 100 pps.

--------------------Figure 1 about here-----------------

In contrast to previous studies, these results show that phoneme recognition performance asymptotes at fairly low stimulation rates. Lawson et al. (1996) predicted that consonant scores should drop significantly when the stimulation rate is reduced from 2525 pps to 250 pps. In the present results there was no significant change in performance between stimulation rates of 500 and 150 pps/channel. However, subjects did report a difference in speech quality when the stimulation rate was 200 pps or 150 pps, noting that speech sounded increasingly "machine-like" or "weird". Phoneme recognition performance deteriorated only when the stimulation rate was reduced to less than 150 pps.

Due to the hardware limitations of the speech processor, the maximum stimulation rate possible in the present study was 500 pps/channel. The present data are not able to exclude the possibility of improvements in performance at higher stimulation rates. Although stochastic neural firing may occur when the speech processor delivers very high stimulation rates, it is not clear if this improves speech performance. Previous speech recognition data obtained at higher rates showed essentially no improvement over lower rates (Lawson et al., 1996; Brill et al., 1997) and in some cases even a decrease in performance at higher rates (Vandali et al., 1998) .

Figure 2 presents the information received on the consonant features of voicing, manner and place of articulation (Miller and Nicely, 1955). Reception of voicing and place cues increased with stimulation rate up to 150 Hz and then did not increase as the stimulation rate was increased above 150 Hz. Information received on manner was essentially unaffected by changes in stimulation rate. At stimulation rates below 150 Hz, the stimulation rate would produce a strong pitch itself, and would not be able to represent voice fundamental frequencies above 75 Hz. Also at stimulation rates below 150 Hz the pitch of the carrier may have interfered with the correct identification of the place of articulation, i.e. the rate pitch of the carrier may have confounded the place pitch of the electrodes. Cues to consonant manner are primarily contained in low-frequency envelope properties which were not affected by stimulation rate even down to 50 Hz.

----------------------Insert Figure 2 about here-------------------

Phoneme recognition was significantly poorer when the stimulation rate was 100 pps-per-channel or lower. One possible explanation is that performance drop is due to the lower temporal resolution transmitted by the envelope filters. As described in the signal processing section, the cutoff frequency of envelope filters was reduced as the stimulation rate was decreased in order to remove any aliasing effects. For example, an envelope filter with a 60 Hz cutoff frequency was used when the stimulation rate was 150 pps, while 20 Hz was used for the stimulation rate of 50 pps. It is possible that the loss of temporal envelope information within each channel was due to the envelope filter rather than the lower stimulation rate *per se*. It is important to distinguish if the limitation is due to the front-end speech processing or to the limitations of perception in electrical stimulation. To determine if the low-pass envelope filter cutoff frequency was the factor limiting speech recognition rather than the stimulation rate, an additional experiment was conducted.

II. EXPERIMENT 2: EFFECT OF CUTOFF FREQUENCY OF ENVELOPE FILTERS

Experiment 2 measured phoneme recognition in conditions that held the stimulation rate per channel constant while reducing the envelope filter cutoff frequency. Vowel and consonant recognition was measured as a function of cutoff frequency of the envelope filters in both cochlear implant listeners using a custom four-channel CIS speech processor and in normal-hearing subjects listening to speech processed by a comparable four noise-band acoustic processor (Shannon et al., 1995).

A. METHOD

1. Subjects

The three cochlear implant listeners from Experiment 1 and five normal-hearing listeners aged 25 to 35 participated in this experiment. All normal-hearing subjects had thresholds better than 15 dB HL at audiometric test frequencies from 250 to 8000 Hz and all were native speakers of American English.

2. Test materials and Procedures

The same test materials and procedures as Experiment 1 were used in this experiment.

3. Signal processing

In electric hearing, the 4-channel CIS processors were similar to those described in Experiment 1 except that, when extracting the envelope in each band, the cutoff frequency of the envelope filters was varied from 2 Hz to 160 Hz. Seven cutoff frequencies were tested: 2, 5, 10, 40, 80, 120, and 160 Hz. The stimulation rate of pulse trains modulated with envelope information was fixed at 500 pps per channel.

In acoustic hearing, the speech signal was spectrally degraded using a four-band modulated noise processor (Shannon et al., 1995). Envelope extraction in each band was the same as that in electric stimulation, except that the cutoff frequency of envelope filters was varied over a broader range: 2 Hz to 640 Hz. Seven envelope filter cutoff frequencies were tested: 2, 5, 10, 20, 40, 160, and 640 Hz. Instead of modulating a continuous pulse train as in electric stimulation, the extracted envelope waveform in each band in the acoustic processor was used to modulate wideband noise that was subsequently spectrally limited by the same bandpass filter used for the analysis filter band. The outputs from all modulated noise bands were then summed and the processed speech tokens were equated in terms of rms energy. All processed speech stimuli were stored on computer disk and were presented via custom software to a 16 bit D/A converter (TDT DD1) at a 16-kHz sampling rate. A-weighted stimuli were presented to the listeners diotically through Sennheiser HDA200 headphones at 70 dB. Within each test, speech stimuli were presented in random order, and test conditions were pseudo-randomized for each subject.

B. RESULTS AND DISCUSSION

Figure 3 shows the vowel and consonant recognition scores as a function of the cutoff frequency of envelope filters. Figure 3A shows the individual and mean vowel scores from three cochlear implant users and the mean vowel scores from five normal-hearing listeners (no significant effect of subjects [F(4,30)=0.361, p=0.834]). Vowel scores increased from 48% to 65% for normal-hearing listeners and from 33% to 50% for implant listeners when the cutoff frequency was increased from 2 Hz to 10 Hz. No significant improvement in either cochlear implant users or normal-hearing listeners was observed when the cutoff frequency was further increased above 10 Hz. There was a significant effect of the cutoff frequency on vowel recognition for both cochlear implant users [F(6,35)=10.02, p<0.001] and for normal-hearing listeners [F(6,28)=35.41, p<0.001]. Post-hoc Tukey HSD tests indicated that the vowel score was significantly lower when the cutoff frequency was 5 Hz or less. An ANOVA also showed no significant interaction between subjects and cutoff frequency for cochlear implant users [F(12,21)= 0.981, p=0.496] as well as no significant interaction between electric hearing and acoustic hearing on vowel recognition [F(6,63)=0.093, p=0.997].

Figure 3B shows the individual and mean consonant scores from three cochlear implant users and the mean consonant scores from five normal-hearing listeners (no significant effect of subjects [F(4,30)=0.016, p=0.999]). When the envelope cutoff frequency was 2 Hz, consonant recognition was nearly at a chance level for cochlear implant users. When the cutoff frequency was increased from 2 to 20 Hz, consonant scores increased dramatically to 65% correct. For normal-hearing listeners, consonant scores also increased dramatically from 17% to 84% when the cutoff frequency was increased from 2 Hz to 20 Hz. Similar to vowel recognition, no significant improvement in either cochlear implant users or normal-hearing listeners was observed when the cutoff frequency was further increased. An analysis of variance showed a significant effect of the cutoff frequency on consonant recognition for both cochlear implant users [F(6,35)=83.56, p<0.001] and for normal-hearing listeners [F(6,28)=248.41, p<0.001]. Post-hoc Tukey HSD tests indicate that the consonant score was significantly lower when the cutoff frequency was 10 Hz or less. Further ANOVA tests showed no significant interaction between subjects and cutoff frequency for cochlear implant users [F(12,21)=1.874, p=0.100] as well as no significant interaction between electric hearing and acoustic hearing on consonant recognition [F(6,63)=1.28, p=0.281].

The fact that implant and normal hearing listeners showed the same pattern of performance as a function of the envelope cutoff frequency suggests that the effect is not specific to electrical stimulation and likely represents a limitation in speech recognition that is common to electric and acoustic hearing. In previous psychophysical studies (e.g., Shannon, 1993) relatively little difference was observed between acoustic and electric hearing on measures of basic temporal processing. The present result extends this finding to the speech envelope domain.

Another interesting observation is that the effect of the envelope cutoff frequency on speech recognition was significantly different for vowels and consonants. Vowel recognition was only moderately reduced as the cutoff frequency was reduced to 2 Hz, where cochlear implant (CI) listeners still scored 35% percent correct and normal-hearing (NH) listeners scored 48% correct. However, consonant recognition scores dropped dramatically from 70% to near-chance level as the cutoff frequency was reduced from 20 Hz to 2 Hz. This result indicates that both CI and NH listeners relied upon temporal envelope cues more for consonant recognition than for vowel recognition.

One might think that high-frequency temporal envelope information would become less important as the number of spectral channels increases, i.e. temporal cues that are important when there are few spectral cues might be less important when more spectral cues are available. Temporal envelope information below 20 Hz can provide sufficient information for modest levels of consonant recognition, even in the complete absence of spectral information (Rosen, 1992;Van Tasell et al., 1987,1992). The present results replicate Shannon et al.'s (1995) finding of no significant performance difference when envelope information above 16-20 Hz was included in 4-band speech processors. Shannon et al. (1995) specifically found no difference in the effect of the low-pass frequency cutoff of the envelope filter between 1, 2, 3, or 4-band processors, indicating that temporal envelope information below 20 Hz is important, regardless of the degree of spectral resolution.

------------------Figure 3 about here------------------

Note that the mean consonant score was about 4%-5% lower for the processor with the 20-Hz cutoff frequency when compared to the processor with the 80-Hz cutoff frequency in both normal-hearing listeners and cochlear implant listeners. Due to the overall performance variation across subjects, this difference was not statistically significant when combining the data from all subjects.

Figure 4 shows the analysis of consonant information received on the production-based features of voicing, manner, and place (Miller and Nicely, 1955) from the data of Experiment 2`. Information received for all three speech features increased as the envelope cutoff frequency increased up to 20 Hz. For normal-hearing listeners there was no change in information received on manner and place as the envelope cutoff frequency was increased above 20 Hz, but voicing information increased slightly up to 160 Hz. In implant listeners a similar pattern of results was observed on voicing and manner, although the level of information received was about 20 percentage points lower than the normal hearing results. However, on place of articulation, information received continued to increase beyond 20 Hz for the implant listeners.

------------------Figure 4 about here------------------

It is not too surprising that higher accuracy of information transmission for voicing was received in speech processors with higher cutoff frequency of envelope filters, presumably because the F_0 information could be at least partly transmitted by the temporal modulation when the cutoff frequency of envelope filters was increased above 80 Hz. However, it is quite puzzling that the information received for place continued to increase in cochlear implant listeners and not in normal-hearing listeners. Whether this difference is due to the high variability of the results across cochlear implant listeners or is an important perceptual difference between acoustic simulation and electric stimulation remains to be answered. One possibility is that rapid spectral transitions in consonants are important for correct identification. In normal-hearing listeners, in whom the noise-band carriers are in the correct tonotopic location, the spectro-temporal transitions may not be as critical for consonant identity. But the temporal information may be more important in implant listeners, in whom the spectral cues are not properly represented by the electrode locations.

III. GENERAL DISCUSSION

The data from the present study show no significant differences in consonant or vowel recognition performance for experimental speech processors with stimulation rates ranging from 150 to 500 pps/channel. This suggests that higher stimulation rates may not provide a substantial benefit to speech recognition in electric hearing as expected.

The results from the present study do show significantly lower scores in vowel and consonant recognition for speech processors with stimulation rates of 50 and 100 pps than for processors with higher stimulation rates. The performance drop in speech recognition could arise either from information reduction in the processor (processor domain) or from a perceptual limitation inherent to implants (perceptual domain). One way in which information is clearly reduced is when the cutoff frequency of envelope filters is decreased. In Experiment 1, as the stimulation rate was reduced from 150 pps/channel to 50 pps/channel, the cutoff frequency of envelope filters decreased from 60 Hz to 20 Hz to avoid aliasing effects. The data from the experiment 2 showed that there were no significant improvements in vowel or consonant recognition when the cutoff frequency of envelope filters was increased above 20 Hz. This implies that the observed performance drop at low stimulation rates in Experiment 1 was not caused by the information reduction in the speech processor.

Another possibility is that the performance drop in speech recognition at low stimulation rates was caused by a perceptual limitation of the implant listeners. One explanation is that pulse trains with lower rates produce their own distinctive pitch sensation that may interfere with the envelope information that is being transmitted. Rate pitch produced by an electrical pulse train is strongest in the 50-150 Hz range and becomes weaker as the pulse rate increases above 150 Hz, so that the pitch sensation generally saturates for pulse rates above 300 Hz (Shannon, 1983; Tong et al., 1983; Townsend et al., 1987; McKay et al., 1996), although a few "star" implant patients may be able to detect pitch differences up to stimulation pulse rates as high as 1 kHz (Pijl and Schwarz, 1993; Wilson et al., 1998). A stimulation pulse train that produces its own strong pitch may interfere with the envelope information that is being transmitted. Even though there was no difference in phoneme recognition for stimulation rates of 150-300 Hz in the present study, patients did report that these processors produced strange and unpleasant sound quality, possibly due to the interfering pitch of the carrier signal.

Both the saturation of rate pitch with stimulation rate and the reduction in speech recognition at low rates might be explained by a relatively slow central temporal integration. Moore et al. (1996) found that the best equivalent rectangular duration for the central temporal integrator is about 7 ms. In the case of stimulation rate, at 150 pps, the pulse separation between two successive pulses is 6.67 ms, while at 100 pps the pulse separation is to 10 ms. Poorer central integration at 100 Hz may cause the envelope information to "break up" and not provide a adequate carrier for the speech envelope information, a phenomenon that might be comparable to exceeding the flicker fusion threshold in vision. The ability to compare successive time frames of speech patterns may influence speech perception. If consecutive frames occur quickly enough, close comparison between two frames provides enough information for speech temporal pattern recognition. If consecutive frames occur too far apart in time, the perceptual "image" breaks up in time, resulting in poorer speech recognition. However, the direct relationship between the central temporal integration and speech pattern recognition is still not clear. Implant patients who perceive rate pitch up to high stimulation rates may have shorter central temporal integration times and thus may be able to use temporal information in speech at stimulation rates higher than 150 Hz.

IV. SUMMARY AND CONCLUSIONS

High stimulation rates may not provide the substantial benefit in electric hearing as previously hypothesized. Results from the present experiments suggest that the stimulation rate can be as low as 150 pps per channel without significantly reducing speech recognition. The cutoff frequencies of envelope filters for different stimulation rates were not sufficient to explain the effect of the stimulation rate on speech recognition, suggesting that the limiting factor was perceptual rather than a limitation inherent to the signal processing. The results are consistent with previous psychophysical results on normal-hearing listeners, which show a 7 ms equivalent rectangular duration of the central temporal integrator.

ACKNOWLEDGEMENTS

We appreciate the efforts of all subjects, especially the three implant listeners for their tireless and heroic participation in our experiments. We thank Professor James Hillenbrand for allowing us to use the multi-talker vowel test materials. The assistance of Xiaosong Wang in data collection and John Galvin III in editorial comments is also greatly appreciated. Work funded by NIH grants DC1526 and DC3861.

REFERENCES

Brill, S.M., Gstöttner, W., Helms, J., Ilberg, C.v., Baumgartner, W., Müller, J., and Kiefer, J. (**1997**). "Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+," Am. J. Oto., 18(Suppl.), S104-S106.

Fishman K., Shannon R.V. and Slattery W.H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," J. Speech Hear Res. **40**, 1201-1215.

Fu, Q.-J., and Shannon, R.V. (**1998**). "Effects of amplitude nonlinearity on speech recognition by cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **104**, 2570-2577.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1994**). "Acoustic characteristics of American english vowels," J. Acoust. Soc. Am. **97**, 3099-3111.

Lawson, D.T., Wilson, B.S., Zerbi, M., and Finley, C.C. (**1996**). Speech processors for auditory prosthesis, NIH project N01-DC-5-2103, Third Quarterly progress report.

McKay, C.M., McDermott, H.J., and Clark, G.M. (**1996**). The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees, J. Acoust. Soc. Am., **99**, 1079-1090.

Miller, G. and Nicely, P. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338-352.

Moore, B.J., Peters, R.W., and Glasberg, B.R. (**1996**). "Detection of decrements and increments in sinusoids at high overall levels," J. Acoust. Soc. Am. **99**, 3669-3677.

Pijl, S. and Schwarz, D.W.F. (**1995**). Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes, J. Acoust. Soc. Am., **98**, 886-895.

Rosen, S. (**1992**). "Temporal information in speech: acoustic, auditory and linguistics aspects," Phil. Trans R. Soc. Lond. B. **336**, 367-373.

Shannon, R.V. (**1983**). Multichannel electrical stimulation of the auditory nerve in man: I. Basic psychophysics, Hearing Research, **11**, 157-189.

Shannon, R.V. (**1993**). Psychophysics of electrical stimulation, in Cochlear Implants: Audiological Foundations, R.S. Tyler (Ed.), Singular Pub. Grp., San Diego, pp. 357-388.

Shannon, R.V., Adams, D.D., Ferrel, R.L., Palumbo, R.L., and Grantgenett, M. (**1990**). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," J. Acoust. Soc. Am. **87**, 905-907.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech Recognition with Primarily Temporal Cues," Science **270**, 303-304.

Tong, Y.C., Blamey, P.J., Dowell, R.C., and Clark, G.M. (**1983**). Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant, J. Acoust. Soc. Amer., **74**, 73-80.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R.L. (**1987**). "Pitch perception by cochlear implant subjects". J. Acoust. Soc. Am., **82**, 106-115.

Vandali, A.E., Grayden, D.B., Whitford, L.A., Plant, K.L. and Clark, G.M. (**1998**). "An analysis of high rate speech processing strategies using the Nucleus 24 Cochlear Implant," presented at the meeting Cochlear Implants in Children, June 4-7, 1998, Iowa City, Iowa.

Van Tasell, D.J., Greenfield, D.G., Logemann, J.J., and Nelson, D.A. (**1992**). "Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants," J. Acoust. Soc. Amer. **92**, 1247-1257.

Van Tasell, D.J., Soli, S.D., Kirby, V.M., and Widin, G.P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Amer., **82**, 1152-1161.

Wilson, B., Finley, C., Zerbi, M., Lawson, D. and van den Honert, C. (**1997**). Speech processors for auditory prosthesis, NIH project N01-DC-5-2103, Seventh Quarterly progress report.

Wilson, B.S., Finley, C.C. Lawson, D., and Zerbi, M. (**1998**). Temporal representations with cochlear implants, American Journal of Otollogy, **18**(Suppl.), S30-S34.

Wilson. B. S., Finley, C. C, Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (**1991**). "New levels of speech recognition with cochlear implants," Nature, **352**, 236-238.

Wilson, B.S., Finley, C.C. Lawson, D., and Zerbi, M. (**1998**). "Temporal representations with cochlear implants," American Journal of Otology **18**(Suppl.), S30-S34.

TABLE

Table 1: Subject information for three Nucleus-22 cochlear implant listeners who participated in the present study. Frequency table refers to the frequency allocation used by the listener in their everyday processor. Frequency table 7 has a frequency range of 120 to 8658 Hz while frequency table 9 has a range of 150 to 10823 Hz. Frequency table 9 is intended to be an approximate tonotopic map to the electrode locations for a full electrode insertion. Insertion depth is reported as the number of stiffening rings outside the round window from the surgical report. A full insertion would be 0 rings out. Sentence and vowel scores are from subjects' own Spectra 22 processor.

Subject	Age	Gender	Cause of Deafness	Duration of use	Insertion Depth	Freq. Table	Score (CUNY)	Score (Vowel)
N3	55	M	Trauma	6 years	3 rings out	7	79.4%	58.1%
N4	39	M	Trauma	4 years	4 rings out	9	99.0%	74.2%
N7	54	M	Unknown	4 years	0 rings out	9	99.0%	65.6%

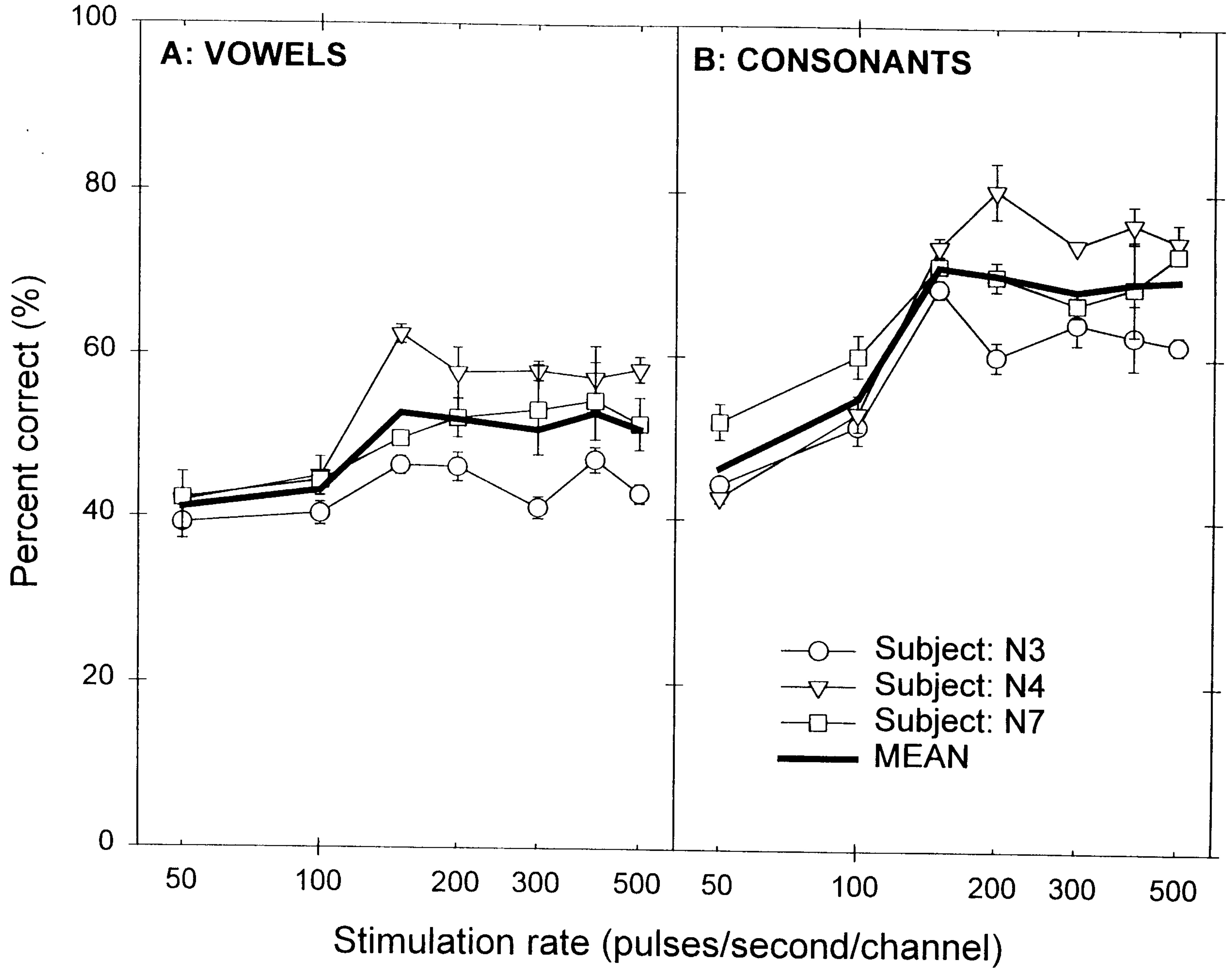
FIGURE CAPTIONS

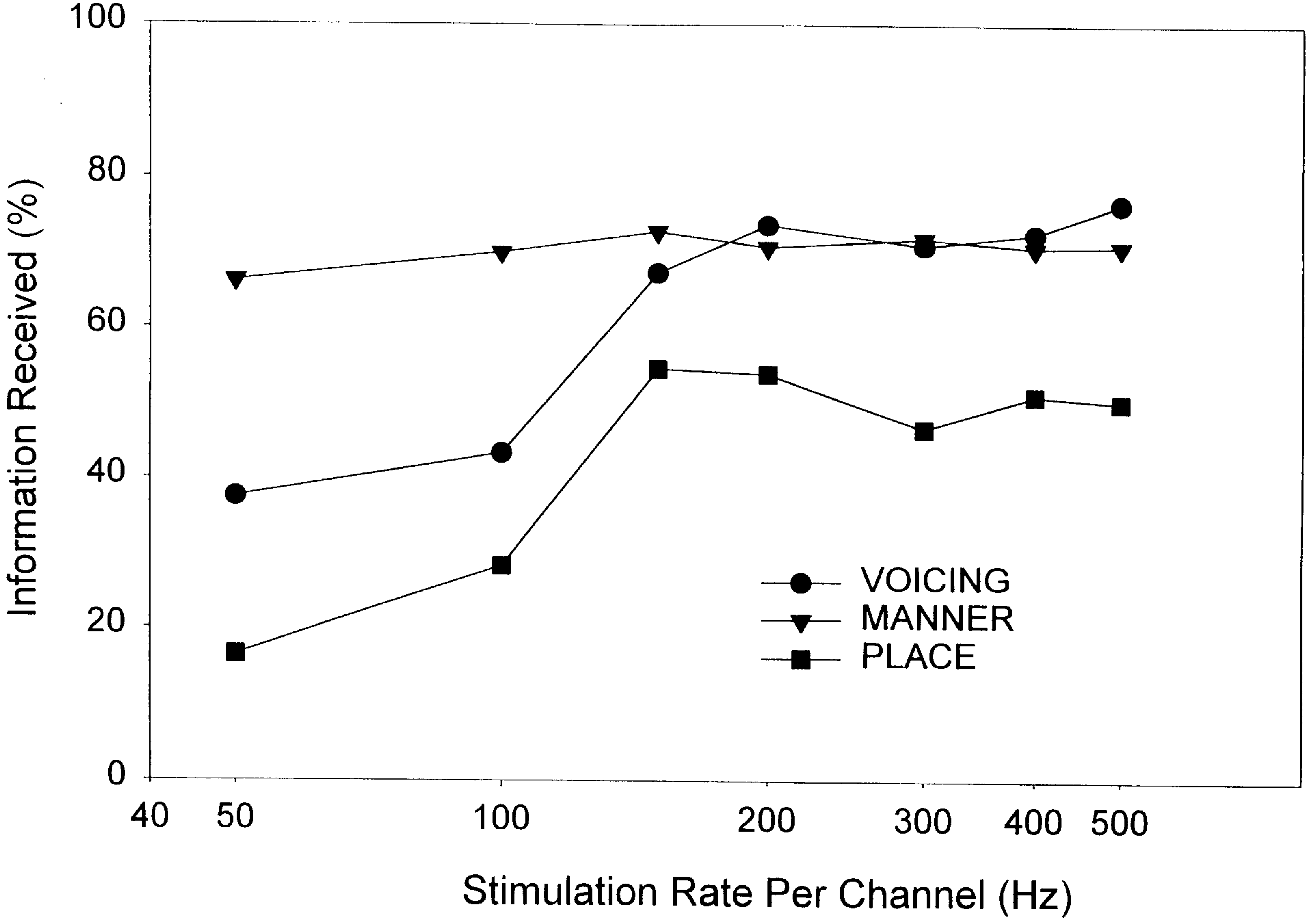
FIGURE 1: Vowel and consonant recognition as a function of the stimulation rate for three Nucleus-22 cochlear implant users. (A) Vowel recognition; (B) Consonant recognition. The cutoff frequency on the low-pass envelope filter was fixed at 40% of the stimulation rate. Error bars represent +/- standard deviation.

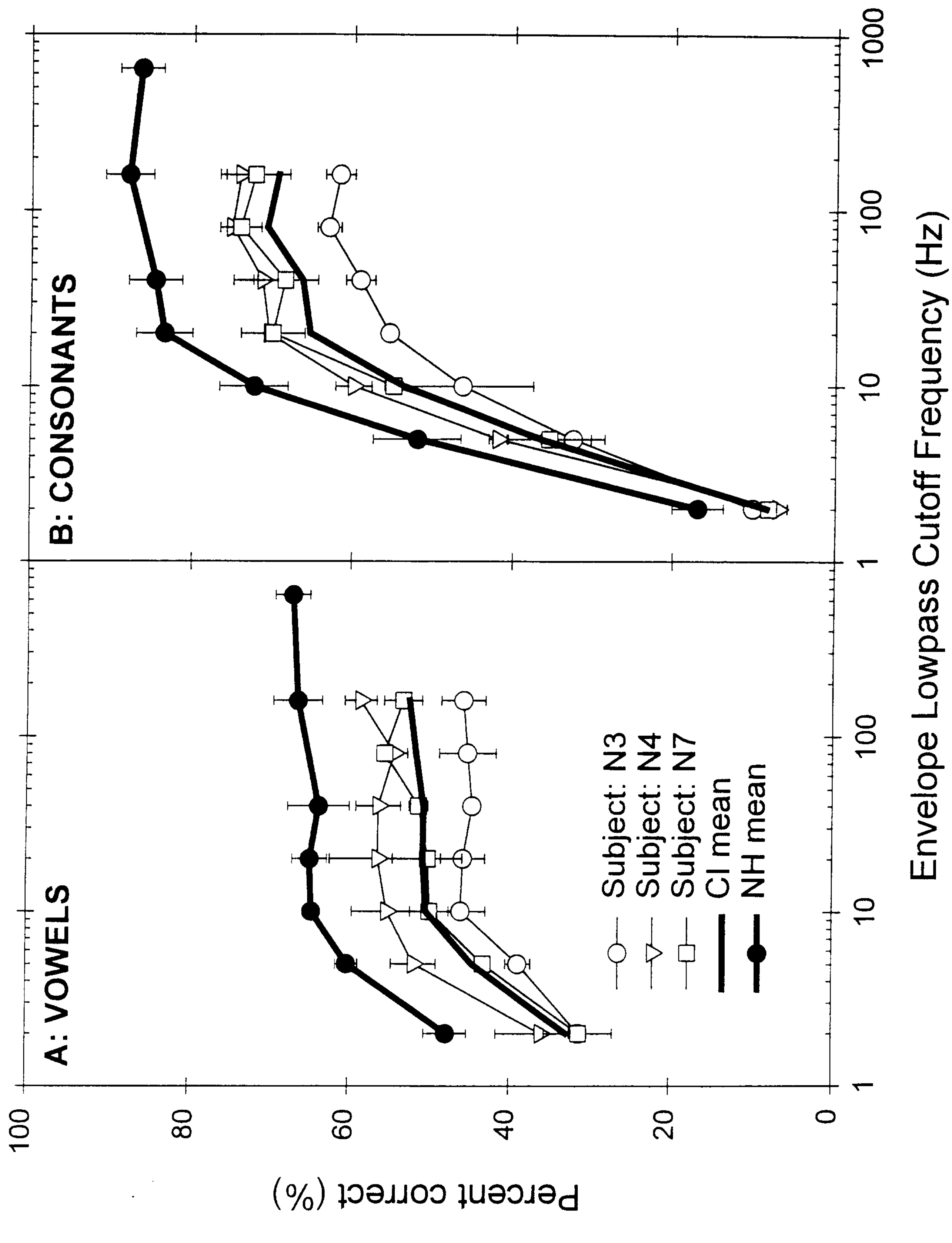
FIGURE 2: Consonant information received on the production based features of voicing, manner and place of articulation as a function of stimulation pulse rate.

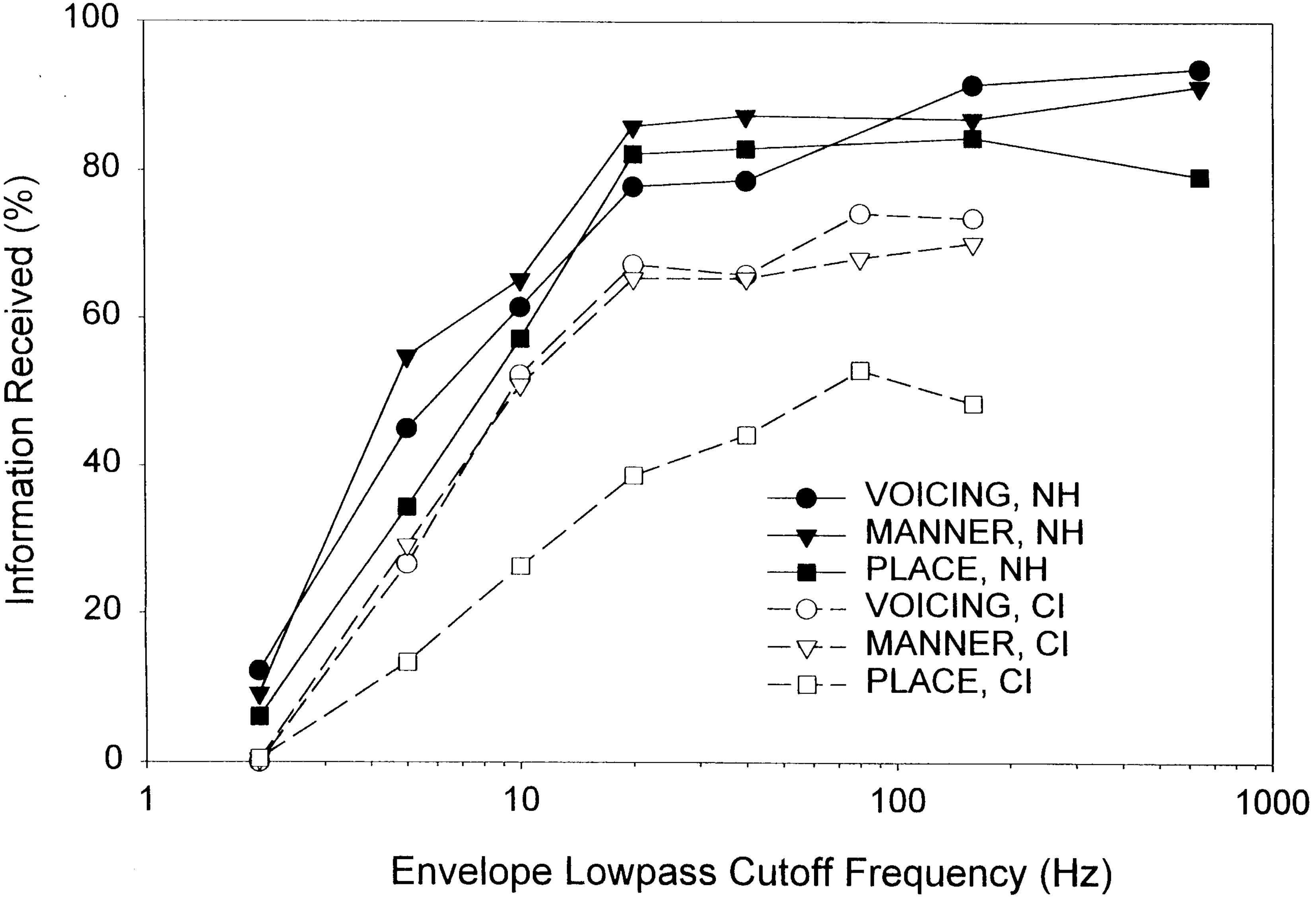
FIGURE 3: Vowel and consonant recognition as a function of the cutoff frequency of the low-pass envelope filters for three Nucleus-22 cochlear implant users and five normal-hearing subjects. (A) Vowel recognition; (B) Consonant recognition. Stimulation rate was fixed at 500 pps/electrode for the implant subjects. Noise bands were used as carriers for the NH listeners. Error bars represent +/- standard deviation.

FIGURE 4: The received information for consonants on voicing (circles), manner (inverted triangles), and place (squares) as a function of the cutoff frequency of the low-pass envelope filters for three Nucleus-22 cochlear implant users (open symbols) and five normal-hearing subjects (filled symbols).


A: VOWELS
B: CONSONANTS
Percent correct (%)
0
20
40
60
80
100
50
100
200
300
500
50
100
200
300
500
Stimulation rate (pulses/second/channel)
Subject: N3
Subject: N4
Subject: N7
MEAN


100
80
60
40
20
0
Information Received (%)
40
50
100
200
300
400
500
Stimulation Rate Per Channel (Hz)
● VOICING
▼ MANNER
■ PLACE


A: VOWELS
B: CONSONANTS
Percent correct (%)
0
20
40
60
80
100
1
10
100
1000
Envelope Lowpass Cutoff Frequency (Hz)
Subject: N3
Subject: N4
Subject: N7
CI mean
NH mean


Information Received (%)
0
20
40
60
80
100
1
10
100
1000
Envelope Lowpass Cutoff Frequency (Hz)
VOICING, NH
MANNER, NH
PLACE, NH
VOICING, CI
MANNER, CI
PLACE, CI